

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:03:43 ; Search time 95.9055 Seconds
(without alignments)
247.473 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 842883

Minimum DB seq length: 0
Maximum DB seq length: 85

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	423	100.0	84	3	AAB18942	Aab18942 Peptide d
2	386	91.3	84	3	AAB18938	Aab18938 Peptide d
3	212	50.1	43	3	AAB18941	Aab18941 Peptide d
4	205	48.5	43	3	AAB18937	Aab18937 Peptide d
5	191	45.2	80	3	AAB18962	Aab18962 Peptide d
6	191	45.2	80	3	AAB18954	Aab18954 Peptide d
7	189	44.7	82	3	AAB18950	Aab18950 Peptide d
8	186	44.0	82	3	AAB18946	Aab18946 Peptide d
9	179	42.3	80	3	AAB18958	Aab18958 Peptide d

10	169	40.0	43	3	AAB18949	Aab18949	Peptide d
11	162	38.3	43	3	AAB18957	Aab18957	Peptide d
12	161	38.1	43	3	AAB18945	Aab18945	Peptide d
13	159	37.6	43	3	AAB18953	Aab18953	Peptide d
14	159	37.6	43	3	AAB18961	Aab18961	Peptide d
15	65.5	15.5	47	4	AAO02215	Aao02215	Human pol
16	61	14.4	77	4	AAO13559	Aao13559	Human pol
17	59.5	14.1	76	4	AAO09950	Aao09950	Human pol
18	58.5	13.8	51	4	ABG07920	Abg07920	Novel hum
19	58.5	13.8	78	4	AAO11222	Aao11222	Human pol
20	56.5	13.4	48	4	AAO04576	Aao04576	Human pol
21	56	13.2	72	4	AAU30892	Aau30892	Novel hum
22	56	13.2	79	5	AAU91124	Aau91124	Human sec
23	56	13.2	79	5	ABG65212	Abg65212	Human alb
24	55.5	13.1	49	4	AAO00883	Aao00883	Human pol
25	55	13.0	73	4	AAO06915	Aao06915	Human pol
26	54.5	12.9	53	4	ABB11592	Abb11592	Human PRO
27	54.5	12.9	57	4	AAO02320	Aao02320	Human pol
28	54.5	12.9	60	3	AAB53484	Aab53484	Human col
29	54.5	12.9	72	3	AAG00338	Aag00338	Human sec
30	54	12.8	51	4	AAO10414	Aao10414	Human pol
31	53	12.5	49	4	AAO13029	Aao13029	Human pol
32	53	12.5	70	4	AAO00993	Aao00993	Human pol
33	53	12.5	75	4	AAU49021	Aau49021	Propionib
34	53	12.5	75	6	ABM45540	Abm45540	Propionib
35	52.5	12.4	52	4	AAO05395	Aao05395	Human pol
36	52.5	12.4	67	4	AAO10683	Aao10683	Human pol
37	52.5	12.4	75	4	AAO09630	Aao09630	Human pol
38	52	12.3	49	4	AAO05487	Aao05487	Human pol
39	52	12.3	53	4	AAU20221	Aau20221	Human nov
40	52	12.3	70	3	AAB28061	Aab28061	Human sec
41	52	12.3	78	4	ABB17410	Abb17410	Human ner
42	51.5	12.2	51	4	AAO09661	Aao09661	Human pol
43	51.5	12.2	64	4	AAO08404	Aao08404	Human pol
44	51.5	12.2	68	4	AAM86399	Aam86399	Human imm
45	51.5	12.2	72	3	AAG03340	Aag03340	Human sec
46	51.5	12.2	72	4	AAB64880	Aab64880	Human sec
47	51.5	12.2	74	4	AAO06480	Aao06480	Human pol
48	51.5	12.2	83	4	AAM94576	Aam94576	Human rep
49	51.5	12.2	83	4	ABG60273	Abg60273	Human ova
50	51.5	12.2	83	5	ABG61744	Abg61744	Novel ova
51	51	12.1	38	4	AAO04567	Aao04567	Human pol
52	51	12.1	52	4	AAO08958	Aao08958	Human pol
53	51	12.1	67	4	AAU29728	Aau29728	Novel hum
54	51	12.1	72	4	AAM95755	Aam95755	Human rep
55	51	12.1	83	4	AAO09594	Aao09594	Human pol
56	51	12.1	84	4	ABG59890	Abg59890	Human liv
57	51	12.1	84	5	ABG47266	Abg47266	Human pep
58	50.5	11.9	55	3	AAG03473	Aag03473	Human sec
59	50.5	11.9	69	4	AAO08892	Aao08892	Human pol
60	50.5	11.9	71	4	AAU43364	Aau43364	Propionib
61	50.5	11.9	71	6	ABM39883	Abm39883	Propionib
62	50.5	11.9	76	4	AAU59836	Aau59836	Propionib
63	50.5	11.9	76	6	ABM56355	Abm56355	Propionib
64	50.5	11.9	80	6	ABP55915	Abp55915	Manduca s
65	50.5	11.9	81	4	AAO02479	Aao02479	Human pol
66	50.5	11.9	83	5	ABP42262	Abp42262	Human ova

67	50	11.8	48	4	AAO09232	Aao09232	Human	pol
68	50	11.8	53	4	AAM91110	Aam91110	Human	imm
69	50	11.8	60	4	AAU32871	Aau32871	Novel	hum
70	50	11.8	74	4	AAO02678	Aao02678	Human	pol
71	50	11.8	77	4	AAM83827	Aam83827	Human	imm
72	50	11.8	78	4	AAU60414	Aau60414	Propionib	
73	50	11.8	78	6	ABM56933	Abm56933	Propionib	
74	49.5	11.7	45	3	AAB34648	Aab34648	Human	sec
75	49.5	11.7	48	4	AAO05087	Aao05087	Human	pol
76	49.5	11.7	49	4	ABB15311	Abb15311	Human	ner
77	49.5	11.7	50	4	AAO06903	Aao06903	Human	pol
78	49.5	11.7	56	4	AAU51696	Aau51696	Propionib	
79	49.5	11.7	56	6	ABM48215	Abm48215	Propionib	
80	49.5	11.7	57	4	AAO11767	Aao11767	Human	pol
81	49.5	11.7	74	3	AAG00397	Aag00397	Human	sec
82	49.5	11.7	80	5	AAE20824	Aae20824	Human	gen
83	49.5	11.7	80	5	AAE20801	Aae20801	Human	gen
84	49.5	11.7	80	5	ABG64654	Abg64654	Human	alb
85	49.5	11.7	80	5	ABG64656	Abg64656	Human	alb
86	49.5	11.7	85	5	ABP58956	Abp58956	Human	zin
87	49	11.6	30	4	AAE03138	Aae03138	Human	gen
88	49	11.6	50	4	AAO06009	Aao06009	Human	pol
89	49	11.6	54	4	AAO09639	Aao09639	Human	pol
90	49	11.6	67	5	ABP08708	Abp08708	Human	ORF
91	49	11.6	68	3	AAB51821	Aab51821	Human	sec
92	49	11.6	69	5	ABP09062	Abp09062	Human	ORF
93	49	11.6	71	4	AAU45158	Aau45158	Propionib	
94	49	11.6	71	6	ABM41677	Abm41677	Propionib	
95	49	11.6	72	2	AAV74050	Aay74050	Human	pro
96	49	11.6	73	4	AAO07109	Aao07109	Human	pol
97	49	11.6	78	4	AAO09281	Aao09281	Human	pol
98	49	11.6	79	4	AAU41349	Aau41349	Propionib	
99	49	11.6	79	6	ABM37868	Abm37868	Propionib	
100	49	11.6	84	4	ABG11223	Abg11223	Novel	hum

ALIGNMENTS

RESULT 1

AAB18942

ID AAB18942 standard; peptide; 84 AA.

XX

AC AAB18942;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.
 XX
 PF 14-MAR-2000; 2000WO-FR000613.
 XX
 PR 15-MAR-1999; 99FR-00003159.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
 XX
 DR WPI; 2000-587566/55.
 XX
 PT Fragments of Grb family proteins to identify compounds are useful in
 PT treating insulin-associated diseases, particularly diabetes and obesity.
 XX
 PS Claim 2; Page 26; 46pp; French.
 XX
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR
 CC is the actual binding region but its effect is about 10 times greater in
 CC presence of SH2 (which by itself is inactive). Agents that affect binding
 CC between the peptides and the insulin receptor can stimulate or inhibit
 CC tyrosine kinase activity of the receptor. The peptides are used for
 CC screening molecules for ability to treat diseases in which insulin is
 CC implicated. The peptides are used to identify agents that are potentially
 CC useful for treating insulin-associated diseases, particularly diabetes
 CC and obesity but also polycystic ovarian syndrome and syndrome X
 XX
 SQ Sequence 84 AA;

Query Match 100.0%; Score 423; DB 3; Length 84;
 Best Local Similarity 100.0%; Pred. No. 8.6e-47;
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
 Qy 61 GTHGSPTASSQSSATNMAIHRSQP 84
 ||||||||||||||||||||
 Db 61 GTHGSPTASSQSSATNMAIHRSQP 84

RESULT 2

AAB18938

ID AAB18938 standard; peptide; 84 AA.

XX

AC AAB18938;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
 XX
 OS Homo sapiens.
 XX
 PN WO200055634-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 14-MAR-2000; 2000WO-FR000613.
 XX
 PR 15-MAR-1999; 99FR-00003159.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
 XX
 DR WPI; 2000-587566/55.
 XX
 PT Fragments of Grb family proteins to identify compounds are useful in
 PT treating insulin-associated diseases, particularly diabetes and obesity.
 XX
 PS Claim 2; Page 25; 46pp; French.
 XX
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR
 CC is the actual binding region but its effect is about 10 times greater in
 CC presence of SH2 (which by itself is inactive). Agents that affect binding
 CC between the peptides and the insulin receptor can stimulate or inhibit
 CC tyrosine kinase activity of the receptor. The peptides are used for
 CC screening molecules for ability to treat diseases in which insulin is
 CC implicated. The peptides are used to identify agents that are potentially
 CC useful for treating insulin-associated diseases, particularly diabetes
 CC and obesity but also polycystic ovarian syndrome and syndrome X
 XX
 SQ Sequence 43 AA;

Query Match 50.1%; Score 212; DB 3; Length 43;
 Best Local Similarity 100.0%; Pred. No. 8.4e-20;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 55
 ||||||||||||||||||||||||||||||||||||||||
 Db 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

RESULT 4

AAB18937

ID AAB18937 standard; peptide; 43 AA.

XX

AC AAB18937;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Rattus sp.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR000613.
XX
PR 15-MAR-1999; 99FR-00003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity.
XX
PS Claim 2; Page 23; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR
CC is the actual binding region but its effect is about 10 times greater in
CC presence of SH2 (which by itself is inactive). Agents that affect binding
CC between the peptides and the insulin receptor can stimulate or inhibit
CC tyrosine kinase activity of the receptor. The peptides are used for
CC screening molecules for ability to treat diseases in which insulin is
CC implicated. The peptides are used to identify agents that are potentially
CC useful for treating insulin-associated diseases, particularly diabetes
CC and obesity but also polycystic ovarian syndrome and syndrome X
XX
SQ Sequence 43 AA;

Query Match 48.5%; Score 205; DB 3; Length 43;
Best Local Similarity 93.0%; Pred. No. 6.8e-19;
Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 55
| | | | : | | | | | | | | | | | | | | : | | | : | | | | | | | | | | | | | |
Db 1 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKK 43

RESULT 5

AAB18962

ID AAB18962 standard; peptide; 80 AA.

XX

AC AAB18962;

XX

DT 06-AUG-2003 (revised)

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS Mus sp.
XX
PN WO200055634-A1.
XX
PD 21-SEP-2000.
XX
PF 14-MAR-2000; 2000WO-FR000613.
XX
PR 15-MAR-1999; 99FR-00003159.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;
XX
DR WPI; 2000-587566/55.
XX
PT Fragments of Grb family proteins to identify compounds are useful in
PT treating insulin-associated diseases, particularly diabetes and obesity.
XX
PS Claim 2; Page 37; 46pp; French.
XX
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR
CC is the actual binding region but its effect is about 10 times greater in
CC presence of SH2 (which by itself is inactive). Agents that affect binding
CC between the peptides and the insulin receptor can stimulate or inhibit
CC tyrosine kinase activity of the receptor. The peptides are used for
CC screening molecules for ability to treat diseases in which insulin is
CC implicated. The peptides are used to identify agents that are potentially
CC useful for treating insulin-associated diseases, particularly diabetes
CC and obesity but also polycystic ovarian syndrome and syndrome X. (Updated
CC on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 80 AA;

RESULT 6
AAB18954
ID AAB18954 standard; peptide; 80 AA.
XX


```

AC      AAB18954;
XX
XX
DT      08-FEB-2001   (first entry)
XX
DE      Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.
XX
KW      Phosphorylated insulin receptor interacting region; Grb7 family protein;
KW      insulin receptor; tyrosine kinase; insulin; insulin-associated disease;
KW      diabetes; obesity; polycystic ovarian syndrome; syndrome X.
XX
OS      Rattus sp.
XX
PN      WO200055634-A1.
XX
PD      21-SEP-2000.
XX
PF      14-MAR-2000; 2000WO-FR000613.
XX
PR      15-MAR-1999;   99FR-00003159.
XX
PA      (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI      Burnol A,  Perdereau D,  Kasus-Jacobi A,  Bereziat V,  Girard J;
XX
DR      WPI; 2000-587566/55.
XX
PT      Fragments of Grb family proteins to identify compounds are useful in
PT      treating insulin-associated diseases, particularly diabetes and obesity.
XX
PS      Claim 2; Page 32; 46pp; French.
XX
CC      B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC      region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR
CC      is the actual binding region but its effect is about 10 times greater in
CC      presence of SH2 (which by itself is inactive). Agents that affect binding
CC      between the peptides and the insulin receptor can stimulate or inhibit
CC      tyrosine kinase activity of the receptor. The peptides are used for
CC      screening molecules for ability to treat diseases in which insulin is
CC      implicated. The peptides are used to identify agents that are potentially
CC      useful for treating insulin-associated diseases, particularly diabetes
CC      and obesity but also polycystic ovarian syndrome and syndrome X
XX
SQ      Sequence 80 AA;

```

AAB18950

XX

AC

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Best Local Similarity 53.0%; Pred. No. 2.1e-16;

Db

Db || : | : : : ||| : |
60 NILGSQSPLHPSTLSTV-IHRTQ 81

RESULT 8

AAB18946

ID AAB18946 standard; peptide; 82 AA.

XX

AC AAB18946;

XX

DT 06-AUG-2003 (revised)

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR000613.

XX

PR 15-MAR-1999; 99FR-00003159.

XX

PA (CNRS) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in

PT treating insulin-associated diseases, particularly diabetes and obesity.

XX

PS Claim 2; Page 28; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting

CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR

CC is the actual binding region but its effect is about 10 times greater in

CC presence of SH2 (which by itself is inactive). Agents that affect binding

CC between the peptides and the insulin receptor can stimulate or inhibit

CC tyrosine kinase activity of the receptor. The peptides are used for

CC screening molecules for ability to treat diseases in which insulin is

CC implicated. The peptides are used to identify agents that are potentially

CC useful for treating insulin-associated diseases, particularly diabetes

CC and obesity but also polycystic ovarian syndrome and syndrome X. (Updated

CC on 06-AUG-2003 to correct OS field.)

XX

SQ Sequence 82 AA;

Query Match 44.0%; Score 186; DB 3; Length 82;

Best Local Similarity 54.1%; Pred. No. 5.1e-16;

Matches 46; Conservative 6; Mismatches 23; Indels 10; Gaps 3;

Query Match 42.3%; Score 179;. DB 3; Length 80;
 Best Local Similarity 59.2%; Pred. No. 4e-15;
 Matches 42; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

Qy 13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQS 72
 |:|| |:|:||||| ||||| |||||:| |||| | | :| :
 Db 13 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEAQAWRKKTNHRLSL---PMPASGT 69
 Qy 73 SATNMAIHRSQ 83
 | : |||:|
 Db 70 S-LSAAIHRTQ 79

RESULT 10

AAB18949

ID AAB18949 standard; peptide; 43 AA.

XX

AC AAB18949;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR000613.

XX

PR 15-MAR-1999; 99FR-00003159.

XX

PA (CNRS) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in
 PT treating insulin-associated diseases, particularly diabetes and obesity.

XX

PS Claim 2; Page 30; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR
 CC is the actual binding region but its effect is about 10 times greater in
 CC presence of SH2 (which by itself is inactive). Agents that affect binding
 CC between the peptides and the insulin receptor can stimulate or inhibit
 CC tyrosine kinase activity of the receptor. The peptides are used for
 CC screening molecules for ability to treat diseases in which insulin is
 CC implicated. The peptides are used to identify agents that are potentially

CC implicated. The peptides are used to identify agents that are potentially
CC useful for treating insulin-associated diseases, particularly diabetes
CC and obesity but also polycystic ovarian syndrome and syndrome X. (Updated
CC on 06-AUG-2003 to correct OS field.)

XX

SQ Sequence 43 AA;

Query Match 38.1%; Score 161; DB 3; Length 43;
Best Local Similarity 78.0%; Pred. No. 3.5e-13;
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 53

||||:||||||| |||:| | | |:| | |

Db 1 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 41

RESULT 13

AAB18953

ID AAB18953 standard; peptide; 43 AA.

XX

AC AAB18953;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Rattus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR000613.

XX

PR 15-MAR-1999; 99FR-00003159.

XX

PA (CNRS) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in

PT treating insulin-associated diseases, particularly diabetes and obesity.

XX

PS Claim 2; Page 32; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein. PIR
CC is the actual binding region but its effect is about 10 times greater in
CC presence of SH2 (which by itself is inactive). Agents that affect binding
CC between the peptides and the insulin receptor can stimulate or inhibit
CC tyrosine kinase activity of the receptor. The peptides are used for

CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 47 AA;

Query Match 15.5%; Score 65.5; DB 4; Length 47;
Best Local Similarity 38.5%; Pred. No. 1;
Matches 15; Conservative 8; Mismatches 13; Indels 3; Gaps 1;

Qy 48 EGLAWRKKGCLRLGTHGS---PTASSQSSATNMAIHR SQ 83
:|: || | |: : || ||::|| | | | : :
Db 4 DGV PWRNPGSLQPPSPGSSDPPTSASQESGTTGAHHHTR 42

RESULT 16

AAO13559

ID AAO13559 standard; protein; 77 AA.

XX

AC AAO13559;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 27451.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation.

XX

OS Homo sapiens.

XX

PN WO200164835-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US004927.

XX

PR 28-FEB-2000; 2000US-00515126.

PR 18-MAY-2000; 2000US-00577409.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-514838/56.

DR N-PSDB; AAI93490.

XX

PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 20; SEQ ID NO 27451; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 77 AA;

Query Match 14.4%; Score 61; DB 4; Length 77;
Best Local Similarity 38.6%; Pred. No. 8;
Matches 17; Conservative 6; Mismatches 17; Indels 4; Gaps 2;

Qy 40 EALSVAVEEGLAWRKKGCLRLGTHGS---PTASSQSSATNMAIH 80
:: ||| | | | | : ||| : ||: ||| | |
Db 18 QSCSVAQARG-QWYNHGSLQPSTHGASNPPTSASQSVGTTGMSH 60

RESULT 17

AAO09950

ID AAO09950 standard; protein; 76 AA.

XX

AC AAO09950;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 23842.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

XX

OS Homo sapiens.

XX

PN WO200164835-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US004927.

XX

PR 28-FEB-2000; 2000US-00515126.

PR 18-MAY-2000; 2000US-00577409.

XX

PA (HYSE-) HYSEQ INC.

XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI89881.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 XX
 PS Claim 20; SEQ ID NO 23842; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 76 AA;

Query Match 14.1%; Score 59.5; DB 4; Length 76;
 Best Local Similarity 40.5%; Pred. No. 12;
 Matches 15; Conservative 8; Mismatches 13; Indels 1; Gaps 1;

Qy 39 TEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSAT 75
 ||: ||| : |: | | | : ||::|| : |
 Db 22 TESRSVA-QAGVQWXLGSLVPGSRHSPSSASQVAGT 57

RESULT 18

ABG07920

ID ABG07920 standard; protein; 51 AA.

XX

AC ABG07920;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #7911.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS72107.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 38279; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 51 AA;

Query Match 13.8%; Score 58.5; DB 4; Length 51;
 Best Local Similarity 42.5%; Pred. No. 9.3;
 Matches 17; Conservative 4; Mismatches 18; Indels 1; Gaps 1;

Qy 46 VEEGLAWRKKGCLRLGTHGSPTAS-SQSSATNMAIHRSP 84
 || | : |:| | | | | | | : ||:|
 Db 12 VEMGFLHVGQAGLKLPTSGDPPASASQAGITGVSHRAQP 51

RESULT 19
 AAO11222
 ID AAO11222 standard; protein; 78 AA.
 XX
 AC AAO11222;
 XX
 DT 06-NOV-2001 (first entry)

XX
 DE Human polypeptide SEQ ID NO 25114.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US004927.
 XX
 PR 28-FEB-2000; 2000US-00515126.
 PR 18-MAY-2000; 2000US-00577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI91153.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 XX
 PS Claim 20; SEQ ID NO 25114; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 78 AA;

Query Match 13.8%; Score 58.5; DB 4; Length 78;
 Best Local Similarity 36.4%; Pred. No. 17;
 Matches 16; Conservative 9; Mismatches 14; Indels 5; Gaps 2;

Qy 40 EALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQ 83
 | : ||| : | : || || : ||| : | : | | : :
 Db 37 ESRVA-QTGVQWRD----LLGSSNSPTSASXVAGTTGACHHAR 75

RESULT 20

AAO04576
ID AAO04576 standard; protein; 48 AA.
XX
AC AAO04576;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 18468.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI84507.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 20; SEQ ID NO 18468; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 48 AA;

Query Match 13.4%; Score 56.5; DB 4; Length 48;
Best Local Similarity 48.4%; Pred. No. 15;
Matches 15; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

Qy 55 KGCLRLGTHGSPTAS-SQSSATNMAIHRSQP 84
 :| | | | | | | | : | | :
 Db 6 QGSLELLTSGDPPASASQSAEITGVSHRTQP 36

RESULT 21

AAU30892

ID AAU30892 standard; protein; 72 AA.

XX

AC AAU30892;

XX

DT 18-DEC-2001 (first entry)

XX

DE Novel human secreted protein #1383.

XX

KW Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200179449-A2.

XX

PD 25-OCT-2001.

XX

PF 16-APR-2001; 2001WO-US008656.

XX

PR 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-611725/70.

XX

PT Nucleic acids encoding a range of human polypeptides, useful in genetic

PT vaccination, testing and therapy.

XX

PS Claim 20; Page 366; 765pp; English.

XX

CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human

CC secreted proteins of the invention
XX
SQ Sequence 72 AA;

Query Match 13.2%; Score 56; DB 4; Length 72;
Best Local Similarity 27.9%; Pred. No. 32;
Matches 17; Conservative 14; Mismatches 16; Indels 14; Gaps 3;

Qy 8 SQSISPMRSISENSLVAMDFSGQKSRV-----IENPTEALSVA-----VEEGLAWRK 54
| | :|: | | : :: | :| : : ||: | :| ::|||:|:|
Db 11 SSSTNPLSSXXLNKIPSLPSSWEKWXIPPKNNCLSLNPSPP-SLAPSLDDIKEGLSWKK 69

Qy 55 K 55
|
Db 70 K 70

RESULT 22

AAU91124

ID AAU91124 standard; protein; 79 AA.

XX

AC AAU91124;

XX

DT 05-JUN-2002 (first entry)

XX

DE Human secreted protein sequence #44.

XX

KW Human secreted protein; autoimmune disease; hyperproliferative disorder;

KW cardiovascular disorder; cerebrovascular disorder; infection; cancer;

KW nervous system disorder; ocular disorder; epithelial cell proliferation;

KW wound healing; skin aging; sunburn; transplantation; chemotaxis;

KW tissue regeneration; food additive; preservative; cytostatic; cardiant;

KW antiviral; antiallergic; antiinflammatory; antibacterial; antifungal.

XX

OS Homo sapiens.

XX

PN WO200218412-A1.

XX

PD 07-MAR-2002.

XX

PF 17-JAN-2001; 2001WO-US001384.

XX

PR 28-AUG-2000; 2000US-0228086P.

PR 04-JAN-2001; 2001US-0259516P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR;

PI Olsen HS, Moore PA, Wei P, Ebner R, Duan RD, Shi Y, Choi GH;

PI Fiscella M, Ni J;

XX

DR WPI; 2002-269525/31.

DR N-PSDB; ABK54162.

XX

PT Seventeen nucleic acid molecules encoding human secreted proteins, useful
PT in the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases.

XX
PS Claim 11; Page 478-479; 505pp; English.

XX
CC The present invention relates to the isolation of novel human secreted
CC proteins, and the polynucleotide sequences encoding them. The secreted
CC proteins are useful to prevent, treat or ameliorate a medical condition
CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
CC sheep. The secreted proteins are also useful in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Antibodies to
CC the secreted proteins can also be used in alleviating symptoms associated
CC with disorders and in diagnostic immunoassays e.g. radioimmunoassays or
CC enzyme linked immunosorbent assays (ELISA). Disorders which can be
CC diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. cancer, cardiovascular
CC disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral
CC ischaemia, angiogenesis, nervous system disorders e.g. Parkinson's
CC disease, infections caused by bacteria, viruses and fungi and ocular
CC disorders e.g. corneal infection. The polypeptides can also be used to
CC aid wound healing and epithelial cell proliferation, to prevent skin
CC aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. AAU91081-
CC AAU91148 represent human secreted protein sequences

XX
SQ Sequence 79 AA;

Query Match 13.2%; Score 56; DB 5; Length 79;
Best Local Similarity 39.3%; Pred. No. 37;
Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 57 CLRLGTHGSPTASSQSSATNMAIHRSQP 84
|| :| | |: | | : | |||
Db 45 CLSIGQHELPSYSCQPGRKRLLPHHSQP 72

RESULT 23

ABG65212

ID ABG65212 standard; protein; 79 AA.

XX

AC ABG65212;

XX

DT 27-AUG-2002 (first entry)

XX

DE Human albumin fusion protein #1887.

XX

KW Albumin fusion protein; therapeutic protein X; human albumin; HA;

KW human serum albumin; HSA; cancer; reproductive disorder;

KW digestive disorder; immune disorder; endocrine disorder;

KW haematopoietic disorder; neural disorder; connective disorder;

KW cytostatic; antiinfertility; antiinflammatory; antiulcer;

KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;

KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;

KW osteopathic; antiarthritic.

XX

OS Homo sapiens.

OS Synthetic.

XX
 PN WO200177137-A1.
 XX
 PD 18-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-US011988.
 XX
 PR 12-APR-2000; 2000US-0229358P.
 PR 25-APR-2000; 2000US-0199384P.
 PR 21-DEC-2000; 2000US-0256931P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Haseltine WA;
 XX
 DR WPI; 2002-010886/01.
 XX
 PT New fusion protein for treating disease e.g. diabetes comprises an
 PT albumin fused to a therapeutic protein.
 XX
 PS Claim 1; Page 1828-1829; 2102pp; English.
 XX
 CC The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA, also known as human serum
 CC albumin, HSA). The proteins are useful for treating a disease or disorder
 CC that may be modulated by therapeutic protein X. The albumin extends the
 CC shelf-life of protein X, and may increase its biological in vitro/in vivo
 CC activity. The protein is useful for treating and diagnosing disorders
 CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
 CC disease, ulcerative colitis), immune disorders (e.g. acquired
 CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
 CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,
 CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
 CC schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).
 CC ABG63326-ABG65518 represent albumin fusion proteins of the invention
 XX
 SQ Sequence 79 AA;

Query Match 13.2%; Score 56; DB 5; Length 79;
 Best Local Similarity 39.3%; Pred. No. 37;
 Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 57 CLRLGTHGSPTASSQSSATNMAIHRSQP 84
 || : | | | : | | : | |||
 Db 45 CLSIGQHELPSYSCQPGRKRLLPHHSQP 72

RESULT 24
 AAO00883
 ID AAO00883 standard; protein; 49 AA.
 XX
 AC AAO00883;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 14775.
 XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US004927.
 XX
 PR 28-FEB-2000; 2000US-00515126.
 PR 18-MAY-2000; 2000US-00577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI80814.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 XX
 PS Claim 20; SEQ ID NO 14775; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 49 AA;

Query Match 13.1%; Score 55.5; DB 4; Length 49;
 Best Local Similarity 33.3%; Pred. No. 22;
 Matches 13; Conservative 8; Mismatches 15; Indels 3; Gaps 1;

Qy 48 EGLAWRKKGCLRLGTHGS---PTASSQSSATNMAIHRSQ 83
 :|: || | |: : || ||::|| | | ::
 Db 6 DGVFWRNPGSLKPPSPGSSDPPTSASQECGITGAHHHTR 44

RESULT 25
 AAO06915
 ID AAO06915 standard; protein; 73 AA.
 XX

AC AAO06915;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 20807.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US004927.
 XX
 PR 28-FEB-2000; 2000US-00515126.
 PR 18-MAY-2000; 2000US-00577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI86846.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 XX
 PS Claim 20; SEQ ID NO 20807; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 73 AA;

Query Match 13.0%; Score 55; DB 4; Length 73;
 Best Local Similarity 44.4%; Pred. No. 44;
 Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 58 LRLGTHGSPTASSQSSATNMAIHRSQP 84
 |||| | ::||:| | |||
 Db 23 LRLGLSDPPASASESTGTTGMSHCSQP 49

Search completed: July 8, 2004, 08:19:17
Job time : 99.9055 secs

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:16:33 ; Search time 21.1654 Seconds
(without alignments)
204.891 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 257387

Minimum DB seq length: 0
Maximum DB seq length: 85

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	50	11.8	68	3	US-09-100-804-30	Sequence 30, Appl
2	49	11.6	85	4	US-09-621-976-5104	Sequence 5104, Ap
3	48	11.3	76	4	US-09-673-395A-519	Sequence 519, App
4	47.5	11.2	67	4	US-09-252-991A-26568	Sequence 26568, A
5	45.5	10.8	73	4	US-09-331-930A-22	Sequence 22, Appl
6	45	10.6	76	3	US-09-083-521-5	Sequence 5, Appli
7	44.5	10.5	61	4	US-09-149-476-615	Sequence 615, App
8	44.5	10.5	62	4	US-09-252-991A-32126	Sequence 32126, A
9	44	10.4	29	4	US-09-227-357-623	Sequence 623, App
10	44	10.4	47	3	US-08-776-059-18	Sequence 18, Appl
11	44	10.4	63	4	US-09-227-357-611	Sequence 611, App

12	44	10.4	80	3	US-09-081-320-20	Sequence 20, Appl
13	44	10.4	80	4	US-09-574-141A-20	Sequence 20, Appl
14	44	10.4	80	4	US-09-707-780-20	Sequence 20, Appl
15	43.5	10.3	55	4	US-08-630-915A-111	Sequence 111, App
16	43.5	10.3	71	4	US-09-621-976-6399	Sequence 6399, Ap
17	43.5	10.3	74	3	US-09-267-177-12	Sequence 12, Appl
18	43.5	10.3	79	4	US-09-252-991A-27207	Sequence 27207, A
19	43.5	10.3	83	4	US-09-107-532A-4334	Sequence 4334, Ap
20	43	10.2	53	2	US-08-726-306A-144	Sequence 144, App
21	43	10.2	72	4	US-09-543-681A-5442	Sequence 5442, Ap
22	43	10.2	76	4	US-09-252-991A-29326	Sequence 29326, A
23	43	10.2	81	4	US-09-134-000C-5090	Sequence 5090, Ap
24	43	10.2	84	4	US-09-247-155-173	Sequence 173, App
25	42.5	10.0	61	4	US-08-630-915A-208	Sequence 208, App
26	42.5	10.0	61	4	US-09-621-976-4275	Sequence 4275, Ap
27	42.5	10.0	85	4	US-09-252-991A-32597	Sequence 32597, A
28	42	9.9	36	3	US-09-045-764A-12	Sequence 12, Appl
29	42	9.9	40	1	US-07-641-971B-5	Sequence 5, Appli
30	42	9.9	40	1	US-07-781-248A-5	Sequence 5, Appli
31	42	9.9	47	3	US-08-776-059-16	Sequence 16, Appl
32	42	9.9	55	4	US-09-369-247-109	Sequence 109, App
33	42	9.9	61	4	US-09-134-000C-5679	Sequence 5679, Ap
34	42	9.9	68	4	US-09-107-532A-5556	Sequence 5556, Ap
35	42	9.9	78	4	US-09-134-001C-2848	Sequence 2848, Ap
36	42	9.9	78	4	US-09-540-236-2395	Sequence 2395, Ap
37	42	9.9	81	4	US-09-198-452A-1167	Sequence 1167, Ap
38	42	9.9	84	6	5171684-3	Patent No. 5171684
39	41.5	9.8	66	4	US-09-621-976-5606	Sequence 5606, Ap
40	41.5	9.8	71	4	US-09-621-976-5550	Sequence 5550, Ap
41	41.5	9.8	77	4	US-09-252-991A-24817	Sequence 24817, A
42	41.5	9.8	79	4	US-09-621-976-5293	Sequence 5293, Ap
43	41.5	9.8	80	4	US-09-563-997A-45	Sequence 45, Appl
44	41.5	9.8	82	4	US-09-107-532A-6598	Sequence 6598, Ap
45	41.5	9.8	83	4	US-09-621-976-4950	Sequence 4950, Ap
46	41	9.7	51	3	US-08-927-219-49	Sequence 49, Appl
47	41	9.7	61	4	US-09-107-532A-6894	Sequence 6894, Ap
48	41	9.7	68	4	US-09-328-352-6617	Sequence 6617, Ap
49	41	9.7	72	4	US-09-252-991A-29154	Sequence 29154, A
50	41	9.7	76	3	US-09-246-500B-9	Sequence 9, Appli
51	41	9.7	79	4	US-09-621-976-7338	Sequence 7338, Ap
52	41	9.7	85	4	US-09-621-976-4396	Sequence 4396, Ap
53	40.5	9.6	51	4	US-09-023-905A-17	Sequence 17, Appl
54	40.5	9.6	60	4	US-09-621-976-5980	Sequence 5980, Ap
55	40.5	9.6	65	4	US-09-621-976-3950	Sequence 3950, Ap
56	40.5	9.6	66	2	US-08-459-568-52	Sequence 52, Appl
57	40.5	9.6	66	2	US-08-399-411-52	Sequence 52, Appl
58	40.5	9.6	66	3	US-08-516-859A-52	Sequence 52, Appl
59	40.5	9.6	66	4	US-09-586-472-52	Sequence 52, Appl
60	40.5	9.6	66	4	US-09-528-706-52	Sequence 52, Appl
61	40.5	9.6	68	1	US-08-606-789-2	Sequence 2, Appli
62	40.5	9.6	68	1	US-08-606-789-4	Sequence 4, Appli
63	40.5	9.6	68	2	US-09-111-348-2	Sequence 2, Appli
64	40.5	9.6	68	2	US-09-111-348-4	Sequence 4, Appli
65	40.5	9.6	69	5	PCT-US95-06406A-5	Sequence 5, Appli
66	40.5	9.6	71	4	US-09-621-976-5251	Sequence 5251, Ap
67	40.5	9.6	71	4	US-09-621-976-5365	Sequence 5365, Ap
68	40.5	9.6	79	4	US-09-006-428A-14	Sequence 14, Appl

69	40	9.5	14	4	US-08-424-361B-13	Sequence 13, Appl
70	40	9.5	39	2	US-08-117-981-2	Sequence 2, Appli
71	40	9.5	39	2	US-08-477-081-2	Sequence 2, Appli
72	40	9.5	39	2	US-08-477-081-18	Sequence 18, Appl
73	40	9.5	39	5	PCT-US93-02142-2	Sequence 2, Appli
74	40	9.5	41	1	US-08-112-208C-7	Sequence 7, Appli
75	40	9.5	41	1	US-08-248-819A-7	Sequence 7, Appli
76	40	9.5	41	2	US-08-337-646A-7	Sequence 7, Appli
77	40	9.5	41	2	US-08-856-531-7	Sequence 7, Appli
78	40	9.5	41	2	US-08-856-034-7	Sequence 7, Appli
79	40	9.5	41	3	US-08-927-326-7	Sequence 7, Appli
80	40	9.5	41	4	US-09-379-820A-7	Sequence 7, Appli
81	40	9.5	47	4	US-09-227-357-656	Sequence 656, App
82	40	9.5	53	3	US-08-905-223-326	Sequence 326, App
83	40	9.5	55	3	US-09-057-486-1	Sequence 1, Appli
84	40	9.5	56	4	US-09-055-075C-48	Sequence 48, Appl
85	40	9.5	56	4	US-09-919-124-48	Sequence 48, Appl
86	40	9.5	67	4	US-09-489-039A-9141	Sequence 9141, Ap
87	40	9.5	72	4	US-09-227-357-655	Sequence 655, App
88	40	9.5	77	3	US-09-246-500B-6	Sequence 6, Appli
89	40	9.5	78	4	US-09-489-039A-13889	Sequence 13889, A
90	40	9.5	83	4	US-09-621-976-5396	Sequence 5396, Ap
91	39.5	9.3	24	6	5240706-21	Patent No. 5240706
92	39.5	9.3	63	4	US-09-621-976-7245	Sequence 7245, Ap
93	39.5	9.3	70	4	US-09-489-039A-8070	Sequence 8070, Ap
94	39.5	9.3	70	4	US-09-489-039A-8701	Sequence 8701, Ap
95	39.5	9.3	70	4	US-09-489-039A-11761	Sequence 11761, A
96	39.5	9.3	72	4	US-09-252-991A-17145	Sequence 17145, A
97	39.5	9.3	73	2	US-08-530-569B-5	Sequence 5, Appli
98	39.5	9.3	73	4	US-09-331-930A-2	Sequence 2, Appli
99	39.5	9.3	73	4	US-09-331-930A-19	Sequence 19, Appl
100	39.5	9.3	73	4	US-09-331-930A-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1

US-09-100-804-30

; Sequence 30, Application US/09100804

; Patent No. 6066472

; GENERAL INFORMATION:

; APPLICANT: GONEZ, LEONEL JORGE

; APPLICANT: SARAS, JAN

; APPLICANT: CLAESSON-WELSH, LENA

; APPLICANT: HELDIN, CARL-HENRIK

; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL

; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN

; TITLE OF INVENTION: TYROSINE PHOSPHATASES

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

; STREET: 600 ATLANTIC AVENUE

; CITY: BOSTON

; STATE: MASSACHUSETTS

; COUNTRY: USA

; ZIP: 02210

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,804
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/596,291
; FILING DATE: 09-AUG-1996
; APPLICATION NUMBER: US 08/115,573
; FILING DATE: 01-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09943
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: LO461/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-100-804-30

```

```

Query Match          11.8%; Score 50; DB 3; Length 68;
Best Local Similarity 31.4%; Pred. No. 26;
Matches 16; Conservative 11; Mismatches 20; Indels 4; Gaps 1;

```

```

Qy      14 MRSISENSLVAMDFSGQKSRVI----ENPTEALSVAVEEGLAWRKKGCLRL 60
      :: ||::|| | | | : | :      || :| : : | | || |::
Db      16 VKEISQDSLAARDGDIQEGDVVLKINGTVTENMSLTDAKTLIERSKGKLM 66

```

```

RESULT 2
US-09-621-976-5104
; Sequence 5104, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21

```

; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5104
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-5104

Query Match 11.6%; Score 49; DB 4; Length 85;
Best Local Similarity 44.0%; Pred. No. 50;
Matches 11; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 52 WRKKGCLRLGTHGSPTASSQSSATN 76
| | | : | | | : | |
Db 17 WRKETSLSLKTQGHREESEQTGFTN 41

RESULT 3

US-09-673-395A-519
; Sequence 519, Application US/09673395A
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 519
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-673-395A-519

Query Match 11.3%; Score 48; DB 4; Length 76;
Best Local Similarity 27.4%; Pred. No. 58;
Matches 20; Conservative 9; Mismatches 24; Indels 20; Gaps 4;

Qy 12 SPMRSISENSLVAMD-----FSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL 60
: | | | | : : | | : | | : | |
Db 2 TPKRHFSPNQPVTLQTVGVNLEHACWLAGKK-----PDDRSNRPVRE--AW-KELCDRR 52

Qy 61 GTHGSPTASSQSS 73
| | | : | :
Db 53 SWHRKPTAKTSSN 65

RESULT 4

US-09-252-991A-26568
; Sequence 26568, Application US/09252991A
; Patent No. 6551795

```
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26568
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26568
```

```
Query Match          11.2%; Score 47.5; DB 4; Length 67;
Best Local Similarity 34.4%; Pred. No. 56;
Matches 11; Conservative 5; Mismatches 9; Indels 7; Gaps 1;
```

```
Qy      36 ENPTEALSVAVEEGLAWRKKGCLRLGTHGSPT 67
      :|||      :|| :|||:| | | |
Db      10 KSPT-----RKGLPEGRKGCVRAGEHEKAT 34
```

RESULT 5

```
US-09-331-930A-22
; Sequence 22, Application US/09331930A
; Patent No. 6436670
; GENERAL INFORMATION:
; APPLICANT: ZIMMET, PAUL Z.
; APPLICANT: COLLIER, GREGORY
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
; FILE REFERENCE: 22975-20007.00
; CURRENT APPLICATION NUMBER: US/09/331,930A
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/AU98/00902
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: AU PP0117/97
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: AU PP0323/97
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-331-930A-22
```

```
Query Match          10.8%; Score 45.5; DB 4; Length 73;
Best Local Similarity 29.4%; Pred. No. 1.2e+02;
Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;
```

Qy 26 DFIGQKSRVIENPTEALS-----VAVEEGLAWRK 54
 | | : | | : | | : : : | : | | |
 Db 8 DRLGKKVRIKCNPSDTIGDLKKLIAAQTGTRWEK 41

RESULT 6

US-09-083-521-5

; Sequence 5, Application US/09083521
 ; Patent No. 6048970
 ; GENERAL INFORMATION:
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Guegler, Karl J.
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 ; STREET: 3174 PORTER DRIVE
 ; CITY: PALO ALTO
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/083,521
 ; FILING DATE: Herewith
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CERRONE, MICHAEL C.
 ; REGISTRATION NUMBER: 39,132
 ; REFERENCE/DOCKET NUMBER: PF-0527 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (650) 855-0555
 ; TELEFAX: (650) 845-4166
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 76 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GenBank
 ; CLONE: 1216498

US-09-083-521-5

Query Match 10.6%; Score 45; DB 3; Length 76;
 Best Local Similarity 24.4%; Pred. No. 1.5e+02;
 Matches 19; Conservative 10; Mismatches 15; Indels 34; Gaps 4;

Qy 6 CSSQSISPMSRISSENSLVAMDFSGQKS-RVIENPTEALSVAVEEGLAWRKKGCLRLGTHG 64
 | : | : | | | : | | | : : | |
 Db 26 CNQTSVAP-----FSGNQSISAAPNPTNATT-----RSGC----- 55

Qy 65 SPTASSQSSATNMAIHR 82
 :| ||:| :|: |
Db 56 ---SSLQSTAGLLALSLS 70

RESULT 7

US-09-149-476-615

; Sequence 615, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: PZ002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662

; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11

```
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02
```

```
Query Match          10.5%; Score 44.5; DB 4; Length 61;
Best Local Similarity 37.9%; Pred. No. 1.2e+02;
Matches    11; Conservative    4; Mismatches    7; Indels    7; Gaps    1;
```

```
Qy      51 AWRKKGCLRLGTHGSPTASSQSSATNMAI 79
        ||| |      |:|:||||| : :
Db      33 AWRPSG-----GTGTSSSQSSTQSRTL 54
```

RESULT 8

US-09-252-991A-32126

; Sequence 32126, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32126
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32126

Query Match 10.5%; Score 44.5; DB 4; Length 62;
Best Local Similarity 40.0%; Pred. No. 1.3e+02;
Matches 10; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

Qy 32 SRVIENPTE---ALSVAVEEGLAWR 53
|| |:|| |:| : || |:|
Db 32 SRTPEHPTSCACAISYKIFEGFCWK 56

RESULT 9

US-09-227-357-623

; Sequence 623, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: PZ010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08

```

; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 623
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-227-357-623

```

```

Query Match          10.4%;  Score 44;  DB 4;  Length 29;
Best Local Similarity 36.0%;  Pred. No. 48;
Matches      9;  Conservative      4;  Mismatches    12;  Indels      0;  Gaps      0;

```

```

Qy      60 LGTHGSPTASSQSSATNMAIHRSP 84
      ||:  |  :|| : |  | :||
Db      5  LGSSDPPAEASQIAGTAAVSHHAQP 29

```

RESULT 10
 US-08-776-059-18
 ; Sequence 18, Application US/08776059B
 ; Patent No. 6271368
 ; GENERAL INFORMATION:
 ; APPLICANT: LENTZEN, Hans
 ; APPLICANT: ECK, Jurgen
 ; APPLICANT: BAUR, Axel
 ; APPLICANT: ZINKE, Holger
 ; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)
 ; FILE REFERENCE: 674503-2003
 ; CURRENT APPLICATION NUMBER: US/08/776,059B
 ; CURRENT FILING DATE: 1999-06-19
 ; EARLIER APPLICATION NUMBER: PCT/EP96/02273
 ; EARLIER FILING DATE: 1996-06-25
 ; EARLIER APPLICATION NUMBER: 95109949.8
 ; EARLIER FILING DATE: 1995-06-26
 ; NUMBER OF SEQ ID NOS: 56
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 18
 ; LENGTH: 47
 ; TYPE: PRT
 ; ORGANISM: Saponaria officinalis
 US-08-776-059-18

Query Match 10.4%; Score 44; DB 3; Length 47;
 Best Local Similarity 34.8%; Pred. No. 98;
 Matches 8; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Qy 25 MDFSGQKSRVIENPTEALSVAVE 47
 || :|:||:| | :|::
 Db 6 MDAVNKKARVVKNEARFLIIAQ 28

RESULT 11
 US-09-227-357-611
 ; Sequence 611, Application US/09227357
 ; Patent No. 6342581
 ; GENERAL INFORMATION:
 ; APPLICANT: Fischer et al.
 ; TITLE OF INVENTION: 123 Human Secreted Proteins
 ; FILE REFERENCE: PZ010P1
 ; CURRENT APPLICATION NUMBER: US/09/227,357
 ; CURRENT FILING DATE: 1999-01-08
 ; EARLIER APPLICATION NUMBER: PCT/US98/13684
 ; EARLIER FILING DATE: 1998-07-07
 ; EARLIER APPLICATION NUMBER: 60/051,926
 ; EARLIER FILING DATE: 1997-07-08
 ; EARLIER APPLICATION NUMBER: 60/052,793
 ; EARLIER FILING DATE: 1997-07-08
 ; EARLIER APPLICATION NUMBER: 60/051,925
 ; EARLIER FILING DATE: 1997-07-08
 ; EARLIER APPLICATION NUMBER: 60/051,929
 ; EARLIER FILING DATE: 1997-07-08
 ; EARLIER APPLICATION NUMBER: 60/052,803
 ; EARLIER FILING DATE: 1997-07-08
 ; EARLIER APPLICATION NUMBER: 60/052,732

; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
; LENGTH: 63

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-611

Query Match 10.4%; Score 44; DB 4; Length 63;
Best Local Similarity 28.6%; Pred. No. 1.5e+02;
Matches 18; Conservative 10; Mismatches 27; Indels 8; Gaps 2;

Qy 4 SGCSSQS-----ISPMRSISENSLVAMDFSGQKSR--VIENPTEALSVAVEEGLAWRKK 55
| ||: :| :| | : | |||:| : | :| | : ||:
Db 1 SPCSAAECHNLSLLSSCSLVSSNILFSFPFGQKARCCLFLFYFSASHIAHESRVYSKKE 60

Qy 56 GCL 58
||
Db 61 MCL 63

RESULT 12

US-09-081-320-20

; Sequence 20, Application US/09081320
; Patent No. 6093544
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,147
; FILING DATE: 20-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/069,902
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1722
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-081-320-20

Query Match 10.4%; Score 44; DB 3; Length 80;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 13; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Qy 31 KSRVIEN--PTEALSVAVEEGLAWRK 54
:| |||| |:|||: |:| | |
Db 40 ESIVIENCGPSEALAATVKEVLGGLK 65

RESULT 13

US-09-574-141A-20
; Sequence 20, Application US/09574141A
; Patent No. 6395490
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035005
; CURRENT APPLICATION NUMBER: US/09/574,141A
; CURRENT FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/047,147
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 09/081,320
; PRIOR FILING DATE: 1998-05-19
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Rupestris stem pitting associated virus
US-09-574-141A-20

Query Match 10.4%; Score 44; DB 4; Length 80;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 13; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Qy 31 KSRVIEN--PTEALSVAVEEGLAWRK 54
:| |||| |:|||: |:| | |
Db 40 ESIVIENCGPSEALAATVKEVLGGLK 65

RESULT 14

US-09-707-780-20
; Sequence 20, Application US/09707780
; Patent No. 6399308
; GENERAL INFORMATION:


```
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035006
; CURRENT APPLICATION NUMBER: US/09/707,780
; CURRENT FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: 09/081,320
; PRIOR FILING DATE: 1998-05-19
; PRIOR APPLICATION NUMBER: 60/047,147
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Rupestris stem pitting associated virus
US-09-707-780-20
```

```
Query Match          10.4%; Score 44; DB 4; Length 80;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 13; Conservative 4; Mismatches 7; Indels 2; Gaps 1;
```

```
Qy      31 KSRVIEN--PTEALSVAVEEGLAWRK 54
          :| |||| |:|||: |:| | |
Db      40 ESIVIENCGPSEALAA TVKEVLGGLK 65
```

RESULT 15

US-08-630-915A-111

```
; Sequence 111, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: McCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-630-915A-111

```

```

Query Match          10.3%; Score 43.5; DB 4; Length 55;
Best Local Similarity 41.4%; Pred. No. 1.4e+02;
Matches 12; Conservative 8; Mismatches 4; Indels 5; Gaps 2;

```

```

Qy      16 SISENSLIVAMDFS-GQKSRVIENPTEALS 43
      :::: ||||: || ||:|   | | |:
Db      23 TVNKGSLVALGFSDGQEAR----PEEILN 47

```

```

RESULT 16
US-09-621-976-6399
; Sequence 6399, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6399
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6399

```

```

Query Match          10.3%; Score 43.5; DB 4; Length 71;
Best Local Similarity 24.4%; Pred. No. 2.1e+02;
Matches 11; Conservative 8; Mismatches 19; Indels 7; Gaps 1;

```

```

Qy      36 ENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAIH 80
      | |:|   | : | : | | | | : : |
Db      28 EQPSET-----WLSLRRRSCSKRETRSSSTRPKTSATIYLTILH 65

```

RESULT 17

US-09-267-177-12

; Sequence 12, Application US/09267177
 ; Patent No. 6287856
 ; GENERAL INFORMATION:
 ; APPLICANT: Poet, Steven E.
 ; APPLICANT: Ritchie, Branson W.
 ; APPLICANT: Niagro, Frank D.
 ; APPLICANT: Lukert, Phil D.
 ; TITLE OF INVENTION: Vaccines against Circovirus Infections
 ; FILE REFERENCE: 21099.0057
 ; CURRENT APPLICATION NUMBER: US/09/267,177
 ; CURRENT FILING DATE: 1999-03-12
 ; EARLIER APPLICATION NUMBER: 60/077,890
 ; EARLIER FILING DATE: 1998-03-13
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 74
 ; TYPE: PRT
 ; ORGANISM: beak and feather disease virus
 US-09-267-177-12

Query Match 10.3%; Score 43.5; DB 3; Length 74;
 Best Local Similarity 31.1%; Pred. No. 2.2e+02;
 Matches 14; Conservative 5; Mismatches 15; Indels 11; Gaps 2;

Qy 36 ENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAIH 80
 ||| | | | : | | | : : | | :
 Db 30 ENPTS-----PEGLV-----CIGGGAPGGPPDTTNTVATKAPIN 63

RESULT 18

US-09-252-991A-27207

; Sequence 27207, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 27207
 ; LENGTH: 79
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27207

Query Match 10.3%; Score 43.5; DB 4; Length 79;

Best Local Similarity 26.5%; Pred. No. 2.5e+02;
Matches 18; Conservative 6; Mismatches 15; Indels 29; Gaps 3;

```
Qy      43 SVAVEEGLAWR-----KKGCLRLGTHGS-----PTA--SSQSS 73
      |   :|:| |           : || | | ||   | | :::||
Db      1 SPPARKGIAGRRADWSPAGREGPRAGCFRRGRSGSARGRRRRTGQGSRRRPRARRNARSS 60

Qy      74 ATNMAIHR 81
      ||      ||
Db      61 ATGSRRHR 68
```

RESULT 19

US-09-107-532A-4334

; Sequence 4334, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND

THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 4334:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 83 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Enterococcus faecium
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...83
; SEQUENCE DESCRIPTION: SEQ ID NO: 4334:
US-09-107-532A-4334

Query Match 10.3%; Score 43.5; DB 4; Length 83;
Best Local Similarity 31.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 8; Mismatches 22; Indels 3; Gaps 2;

Qy 24 AMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQ 71
|:| || :: : : : | || : | || || | :|
Db 24 ALDGS GDEA--VSSYEGFFADEVHRGL-YHFNGALALGDHGPHTNGNQ 68

RESULT 20

US-08-726-306A-144

; Sequence 144, Application US/08726306A
; Patent No. 5958684
; GENERAL INFORMATION:
; APPLICANT: van Leeuwen, Frederik Willem
; APPLICANT: Burbach, Johannes Peter Henri
; APPLICANT: Grosveld, Franklin G.
; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
; NUMBER OF SEQUENCES: 189
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1 Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,306A
; FILING DATE: 02-Oct-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 95/20080.4
; FILING DATE: 02-Oct-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,832
; FILING DATE: 01-Jan-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 144:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-726-306A-144

Query Match 10.2%; Score 43; DB 2; Length 53;
Best Local Similarity 33.3%; Pred. No. 1.6e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 8; Gaps 1;

Qy 32 SRVIENPTEALSVAVEEGLAWRKKGCL 58
|| || | ||: ||:
Db 6 SRTTRPPT-----SGATWRRPGCI 24

RESULT 21

US-09-543-681A-5442
; Sequence 5442, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5442
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5442

Query Match 10.2%; Score 43; DB 4; Length 72;
Best Local Similarity 46.7%; Pred. No. 2.5e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 43 SVAVEEGLAWRKKGC 57
|::| | | | |
Db 36 SLSIEEGLLWALNKC 50

RESULT 22

US-09-252-991A-29326
; Sequence 29326, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29326
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29326

Query Match 10.2%; Score 43; DB 4; Length 76;
Best Local Similarity 40.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 2; Mismatches 13; Indels 6; Gaps 2;

Qy 4 SGCSSQS--ISPMRSISENSLVAMDFSGQKSRVIE 36
: || || | || | | | | | | |
Db 42 AACSRQSWGIYPM---NQGYKAMPFRGDYHRVVE 72

RESULT 23

US-09-134-000C-5090
; Sequence 5090, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5090
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5090

Query Match 10.2%; Score 43; DB 4; Length 81;
Best Local Similarity 24.4%; Pred. No. 3e+02;
Matches 11; Conservative 11; Mismatches 13; Indels 10; Gaps 2;

Qy 21 SLVAMDFSGQKSRVIEN---PTEALSVAVEEGLA-----WRKK 55
: | : || : | | : : | : || : : : | | :
Db 7 ALEVIDFKSKDKRVNSKKIPPLKAIEVAKRKNVSAATVTRWMKR 51

RESULT 24

US-09-247-155-173
; Sequence 173, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric

```
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 173
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -36..-1
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: -26,-25,-24
; OTHER INFORMATION: Xaa = any one of the twenty amino acids
US-09-247-155-173
```

```
Query Match          10.2%; Score 43; DB 4; Length 84;
Best Local Similarity 36.7%; Pred. No. 3.2e+02;
Matches 11; Conservative 2; Mismatches 13; Indels 4; Gaps 1;
```

```
Qy      37 NPTEALSVAVEEGLAWRKKGCLRLGTHGSP 66
      ||      :||      :|  ||  ||  ||
Db      18 NPDHHSCLAV----SWEAAGCHGAGTQQSP 43
```

RESULT 25

US-08-630-915A-208

```
; Sequence 208, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: McCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
```



```

; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-630-915A-208

```

```

Query Match          10.0%; Score 42.5; DB 4; Length 61;
Best Local Similarity 33.3%; Pred. No. 2.3e+02;
Matches 10; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

```

```

Qy      9 QSISPMRSISENSLVAMDFSGQKSRVIENP 38
        |::| |::| | : |: ||| |
Db      6 QTLYPFSSVTEELNEFE-KGETMEVIEKP 34

```

```

Search completed: July 8, 2004, 08:23:31
Job time : 22.1654 secs

```

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:06:23 ; Search time 17.8583 Seconds
 (without alignments)
 452.456 Million cell updates/sec

Title: US-09-936-697-6
 Perfect score: 423
 Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQSSATNMAIHRSP 84

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 28653

Minimum DB seq length: 0
 Maximum DB seq length: 85

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 100 summaries

Database : PIR 78:*
 1: pir1:*
 2: pir2:*
 3: pir3:*
 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		DB	ID	Description	
	Score	Match Length				
1	54	12.8	80	2	T27603	hypothetical prote
2	51.5	12.2	76	2	E64324	DNA-directed RNA p
3	46.5	11.0	62	2	E64510	hypothetical prote
4	46	10.9	77	2	S24471	gag polyprotein -
5	45.5	10.8	65	2	A82657	hypothetical prote
6	45.5	10.8	73	2	T25763	hypothetical prote
7	45.5	10.8	81	2	JC5345	cddl protein - Clo
8	45	10.6	60	2	A42960	ferredoxin 2[4Fe-4
9	45	10.6	76	2	I53107	CD24 precursor - r
10	45	10.6	77	2	H69420	hydrogenase expres
11	45	10.6	78	2	D81246	hypothetical prote
12	45	10.6	79	2	T17014	metallothionein-li
13	45	10.6	82	2	D69087	hydrogenase expres

14	44.5	10.5	64	2	AB2271	periplasmic mercur
15	44.5	10.5	79	2	S75293	hypothetical prote
16	44	10.4	51	2	D85807	hypothetical prote
17	44	10.4	78	2	B90959	probable phage tai
18	44	10.4	78	2	A75411	hypothetical prote
19	43.5	10.3	83	2	G90914	excisionase [impor
20	43	10.2	82	1	G64370	conserved hypothet
21	42.5	10.0	52	2	D82682	hypothetical prote
22	42.5	10.0	64	2	AC2544	hypothetical prote
23	42.5	10.0	68	2	T00189	hypothetical prote
24	42.5	10.0	74	2	S72807	hypothetical prote
25	42	9.9	36	2	JC2006	differentiation in
26	42	9.9	43	2	S35580	cysteine proteinas
27	42	9.9	45	2	S36709	B7 protein - equin
28	42	9.9	53	2	C82776	hypothetical prote
29	42	9.9	64	2	C96993	hypothetical prote
30	42	9.9	73	2	F69062	molybdenum transpo
31	42	9.9	78	2	D91160	hypothetical prote
32	42	9.9	78	2	C86006	hypothetical prote
33	42	9.9	78	2	E65136	hypothetical 8.7 k
34	42	9.9	79	2	A86517	hypothetical prote
35	42	9.9	79	2	B72106	hypothetical prote
36	42	9.9	81	2	D81565	hypothetical prote
37	41.5	9.8	72	2	A97136	Zn-finger containi
38	41.5	9.8	77	2	B98247	hypothetical prote
39	41.5	9.8	80	2	S00858	T-cell receptor al
40	41	9.7	52	2	A34657	cathepsin E (EC 3.
41	41	9.7	60	2	I68870	interleukin 2 - we
42	41	9.7	62	2	I54512	interleukin 2 - mo
43	41	9.7	65	2	T03652	probable carbonate
44	41	9.7	66	2	F90146	DNA-directed RNA p
45	41	9.7	68	2	D53203	hypothetical prote
46	41	9.7	75	2	S24478	gag polyprotein -
47	40.5	9.6	68	2	B42243	GTP-binding regula
48	40.5	9.6	70	2	E70985	hypothetical prote
49	40.5	9.6	73	2	A60172	proteoglycan core
50	40.5	9.6	77	2	AG3462	hypothetical prote
51	40	9.5	41	2	C47538	bcl-2-associated p
52	40	9.5	51	2	D81039	hypothetical prote
53	40	9.5	63	2	AC2553	hypothetical prote
54	40	9.5	64	2	S22044	gag polyprotein -
55	40	9.5	66	2	A43602	T-cell-stimulating
56	40	9.5	67	2	G97092	endoglucanase (tru
57	40	9.5	70	2	T02541	hypothetical prote
58	40	9.5	72	2	S24477	gag polyprotein -
59	40	9.5	73	2	S24472	gag polyprotein -
60	40	9.5	74	2	S24473	gag polyprotein -
61	40	9.5	75	2	S24475	gag polyprotein -
62	40	9.5	75	2	S24474	gag polyprotein -
63	40	9.5	76	2	T30403	late expression fa
64	40	9.5	77	2	S31660	voltage-dependent
65	40	9.5	78	2	H90114	putative small nuc
66	40	9.5	83	2	A82386	hypothetical prote
67	39.5	9.3	56	2	AB2413	hypothetical prote
68	39.5	9.3	61	2	F86696	4-oxalocrotonate t
69	39.5	9.3	67	2	AF1487	probable transcrip
70	39.5	9.3	67	2	AH1375	repressor protein

71	39.5	9.3	72	2	G97751	hypothetical prote
72	39.5	9.3	75	2	C81951	hypothetical prote
73	39.5	9.3	75	2	A70610	hypothetical prote
74	39.5	9.3	77	2	B95003	hypothetical prote
75	39.5	9.3	77	2	AD1945	hypothetical prote
76	39.5	9.3	81	2	F90454	hypothetical prote
77	39.5	9.3	83	1	W8BPG7	gene 18.7 protein
78	39	9.2	45	1	C64901	ribosomal protein
79	39	9.2	45	2	D90889	30S ribosomal subu
80	39	9.2	45	2	E85728	30S ribosomal subu
81	39	9.2	46	2	PC4162	toxin-co-regulated
82	39	9.2	62	2	T06654	hypothetical prote
83	39	9.2	64	2	A48411	Myf5 homolog - chi
84	39	9.2	69	2	T44123	hypothetical prote
85	39	9.2	72	2	AD3532	hypothetical prote
86	39	9.2	77	2	B83269	hypothetical prote
87	39	9.2	81	1	C70910	hypothetical prote
88	39	9.2	81	2	A97803	hypothetical prote
89	38.5	9.1	48	2	T35253	small hypothetical
90	38.5	9.1	60	2	AC2981	hypothetical prote
91	38.5	9.1	63	2	T31143	hypothetical prote
92	38.5	9.1	64	2	D81172	hypothetical prote
93	38.5	9.1	67	2	T42055	cold shock protein
94	38.5	9.1	67	2	C71854	hypothetical prote
95	38.5	9.1	67	2	AI1126	probable transcrip
96	38.5	9.1	69	2	S70158	hypothetical prote
97	38.5	9.1	72	1	D69550	hypothetical prote
98	38.5	9.1	74	2	A25408	complement C5 - bo
99	38.5	9.1	76	2	B64660	hypothetical prote
100	38.5	9.1	76	2	A82122	hypothetical prote

ALIGNMENTS

RESULT 1

T27603

hypothetical protein ZC477.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T27603

R;Du, Z.

submitted to the EMBL Data Library, November 1995

A;Description: The sequence of *C. elegans* cosmid ZC477.

A;Reference number: Z20392

A;Accession: T27603

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-80 <DUZ>

A;Cross-references: EMBL:U40802; PIDN:AAA81504.1; CESP:ZC477.4

C;Genetics:

A;Gene: CESP:ZC477.4

Query Match 12.8%; Score 54; DB 2; Length 80;

Best Local Similarity 28.4%; Pred. No. 40;

Matches 23; Conservative 9; Mismatches 35; Indels 14; Gaps 2;

Qy 6 CSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGS 65
 || | | | | : | | | : || : | | : | | : |
 Db 4 CSPLKILPGASSSSSSSTA-----SSQIRPPSLSLASLSEELRVEECGSPRVGAKES 56

Qy 66 -----PTASSQSSATNMAI 79
 | | | : :
 Db 57 SFYCTEQPAQSSSYSREDKLCL 77

RESULT 2

E64324

DNA-directed RNA polymerase (EC 2.7.7.6) subunit N - Methanococcus jannaschii

C;Species: Methanococcus jannaschii

C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 23-Apr-1999

C;Accession: E64324

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.;

Woese, C.R.; Venter, J.C.

A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.

A;Reference number: A64300; MUID:96337999; PMID:8688087

A;Accession: E64324

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-76 <BUL>

A;Cross-references: GB:U67475; GB:L77117; NID:g1590930; PID:g1590941;

TIGR:MJ0196; PID:g1510312

C;Genetics:

A;Map position: FOR190573-190803

A;Start codon: GTG

C;Superfamily: DNA-directed RNA polymerase II chain RPB10

C;Keywords: nucleotidyltransferase; transcription

Query Match 12.2%; Score 51.5; DB 2; Length 76;

Best Local Similarity 30.0%; Pred. No. 73;

Matches 15; Conservative 10; Mismatches 18; Indels 7; Gaps 2;

Qy 13 PMRSISENSLVAMDFSGQKSRVI--ENPTEALSVAVEEGLAWRKKGCLRL 60
 | : | | : : | | | : | : | : | : | :
 Db 7 PIRCFSCGNVIAEVFEEYKERILKGENPKDVL-----DDLGIKKYCCRRM 51

RESULT 3

E64510

hypothetical protein MJECL05 - Methanococcus jannaschii plasmid pURB800

C;Species: Methanococcus jannaschii

C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Oct-1999

C;Accession: E64510

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.;

Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.;

Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Query Match 11.0%; Score 46.5; DB 2; Length 62;
Best Local Similarity 28.6%; Pred. No. 2.1e+02;
Matches 12; Conservative 8; Mismatches 21; Indels 1; Gaps 1;

QY 15 RSISENSLVAMDFS-GQKSRVIENPTEALSVAVEEGLAWRKK 55
: ::| | :: | | : | | :|||: | ||
Db 18 KKVAERFLKDLESSQGMWDKEIRERAERAKKOLEEGIEWAKK 59

RESULT 4

Query Match 10.9%; Score 46; DB 2; Length 77;
Best Local Similarity 23.4%; Pred. No. 3.2e+02;
Matches 15; Conservative 7; Mismatches 14; Indels 28; Gaps 2;

Qy 29 GQKSRVI----ENPTEALSVAVEEG-----LAWRKKGCLRL 60
| |::||: | |:: :: | | |||| :
Db 1 GHKARVLAQAMSKATNAATIMMQRGNFRNQKTVKCFNCGKOGHIARNCRAPRKKGCWKC 60

Qy. 61 GTHG 64
 | |
Db 61 GKEG 64

RESULT 5

A82657

hypothetical protein XF1634 [imported] - *Xylella fastidiosa* (strain 9a5c)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C;Accession: A82657

R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: A82657

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-65 <SIM>

A;Cross-references: GB:AE003990; GB:AE003849; NID:g9106683; PIDN:AAF84443.1; GSPDB:GN00128; XFSC:XF1634

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros, M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF1634

Query Match 10.8%; Score 45.5; DB 2; Length 65;
Best Local Similarity 33.3%; Pred. No. 3e+02;
Matches 11; Conservative 6; Mismatches 11; Indels 5; Gaps 1;

Qy 44 VAVEEGLAWRKK-----GCLRLGTHGSPTASSQ 71
: :| || || | | :|||::
Db 16 ICIENTLALRKKNIIYLPNCCTSLEHSAPTATAK 48

RESULT 6

T25763

hypothetical protein F46F11.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T25763

R;Pauley, A.; Gattung, S.

submitted to the EMBL Data Library, February 1997

A;Description: The sequence of *C. elegans* cosmid F46F11.

A;Reference number: Z20083

A;Accession: T25763

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-73 <PAU>

A;Cross-references: EMBL:U88173; PIDN:AAB42266.1; GSPDB:GN00019; CESP:F46F11.4

A;Experimental source: strain Bristol N2; clone F46F11

C;Genetics:

A;Gene: CESP:F46F11.4

A;Map position: 1

A;Introns: 38/2

Query Match 10.8%; Score 45.5; DB 2; Length 73;
Best Local Similarity 29.4%; Pred. No. 3.4e+02;
Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;

Qy 26 DFSGQKSRVIENPTEALS-----VAVEEGLAWRK 54
| |:| |: ||:: : :| :| ||
Db 8 DRLGKKVRIKCNPSDTIGDLKKLIAAQTGTRWEK 41

RESULT 7

JC5345

cddl protein - *Clostridium difficile*

C;Species: *Clostridium difficile*

C;Date: 27-May-1997 #sequence_revision 18-Jul-1997 #text_change 15-Oct-1999

C;Accession: JC5345

R;Braun, V.; Hundsberger, T.; Leukel, P.; Sauerborn, M.; von Eichel-Streiber, C.
Gene 181, 29-38, 1996

A;Title: Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*.

A;Reference number: JC5340; MUID:97128764; PMID:8973304

A;Accession: JC5345

A;Molecule type: DNA

A;Residues: 1-81 <BRA>

A;Cross-references: EMBL:X92982; NID:g1770128; PIDN:CAA63566.1; PID:e212011;
PID:g1770137

A;Experimental source: strain VPI10463

C;Genetics:

A;Gene: cdul

Query Match 10.8%; Score 45.5; DB 2; Length 81;
Best Local Similarity 25.5%; Pred. No. 3.8e+02;
Matches 14; Conservative 8; Mismatches 22; Indels 11; Gaps 1;

Qy 9 QSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTH 63
| | : : | || : : : | | | | : : | | | |
Db 5 QKIPGVGKATEKSLIMLGYYTTIKSLKDANPAQMY-----EKECLMRGQH 48

RESULT 8

A42960

ferredoxin 2[4Fe-4S] - Methanosarcina thermophila

C;Species: Methanosarcina thermophila

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Nov-1998

C;Accession: A42960

R;Clements, A.P.; Ferry, J.G.

J. Bacteriol. 174, 5244-5250, 1992

A;Title: Cloning, nucleotide sequence, and transcriptional analyses of the gene encoding a ferredoxin from Methanosarcina thermophila.

A;Reference number: A42960; MUID:92355496; PMID:1379583

A;Contents: TM-1

A;Accession: A42960

A;Molecule type: DNA

A;Residues: 1-60 <CLE>

A;Note: sequence extracted from NCBI backbone (NCBIN:110322, NCBIP:110324)

C;Genetics:

A;Gene: fdxA

C;Superfamily: ferredoxin 2[4Fe-4S]; ferredoxin 2[4Fe-4S] homology

C;Keywords: 4Fe-4S; electron transfer; iron-sulfur protein; metalloprotein

F;3-59/Domain: ferredoxin 2[4Fe-4S] homology <FER>

F;10,13,16,51/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F;20,41,44,47/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 10.6%; Score 45; DB 2; Length 60;
Best Local Similarity 42.9%; Pred. No. 3.1e+02;
Matches 12; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Qy 24 AMDFSGQKSRVIENPTEALSVAVEEGLA 51
| : || | | | | : || : : | : | : |
Db 7 ADECSGCGSCVDECPSEAITLDEEKGIA 34

RESULT 9

I53107

CD24 precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999

C;Accession: I53107; S25146

R;Shirasawa, T.; Akashi, T.; Sakamoto, K.; Takahashi, H.; Maruyama, N.; Hirokawa, K.

Dev. Dyn. 198, 1-13, 1993

A;Title: Gene expression of CD24 core peptide molecule in developing brain and developing non-neural tissues.

A;Reference number: I53107; MUID:94122434; PMID:8292828

A;Accession: I53107

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-76 <RES>
A;Cross-references: EMBL:Z11663; NID:g55901; PIDN:CAA77731.1; PID:g55902
C;Keywords: phosphatidylinositol linkage

Query Match 10.6%; Score 45; DB 2; Length 76;
Best Local Similarity 24.4%; Pred. No. 4.1e+02;
Matches 19; Conservative 10; Mismatches 15; Indels 34; Gaps 4;

Qy 6 CSSQSISPMRSISENSLVAMDFSGQKS-RVIENPTEALSVAVEEGLAWRKKGCLRLGTHG 64
|: |:| ||| :| ||| | : : ||
Db 26 CNQTSVAP-----FSGNQSISAAPNPTNATT-----RSGC----- 55

Qy 65 SPTASSQSSATNMAIHR 82
:| ||:| :|: |
Db 56 ---SSLQSTAGLLALSLS 70

RESULT 10

H69420

hydrogenase expression/formation protein (hypC) homolog - Archaeoglobus fulgidus

C;Species: Archaeoglobus fulgidus

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 14-Apr-2003

C;Accession: H69420

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: H69420

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-77 <KLE>

A;Cross-references: GB:AE001009; GB:AE000782; NID:g2689332; PIDN:AAB89878.1; PID:g2649207; TIGR:AF1369

C;Superfamily: [NiFe]-hydrogenase maturation chaperone

Query Match 10.6%; Score 45; DB 2; Length 77;
Best Local Similarity 35.1%; Pred. No. 4.1e+02;
Matches 13; Conservative 5; Mismatches 15; Indels 4; Gaps 1;

Qy 22 LVAMDFSGQKSRV----IENPTEALSVAVEEGLAWRK 54
: :|| | | :|| | | :|
Db 16 IAIVDFKGLKKEVRIDLLENPQIGDYVLVHVGMAIQK 52

RESULT 11

D81246
hypothetical protein NMB0016 [imported] - *Neisseria meningitidis* (strain MC58 serogroup B)
C;Species: *Neisseria meningitidis*
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: D81246
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Ketchum, K.A.; Hood, D.W.; Peden, J.F.; Dodson, R.J.; Nelson, W.C.; Gwinn, M.L.; DeBoy, R.; Peterson, J.D.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Mason, T.; Ciecko, A.; Parksey, D.S.; Blair, E.; Cittone, H.; Clark, E.B.; Cotton, M.D.; Utterback, T.R.; Khouri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiagnani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter, J.C.
A;Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: D81246
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-78 <TET>
A;Cross-references: GB:AE002360; GB:AE002098; NID:g7225241; PIDN:AAF40495.1; PID:g7225242; GSPDB:GN00119; TIGR:NMB0016
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0016

Query Match 10.6%; Score 45; DB 2; Length 78;
Best Local Similarity 41.7%; Pred. No. 4.2e+02;
Matches 10; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

Qy 39 TEALSVAVEEGLAWR--KKGCLRL 60
|||::|:||:| |||
Db 21 TEWLPMSLRTGILWRFERKVCLEL 44

RESULT 12
T17014
metallothionein-like protein AMT1 - apple tree
C;Species: *Malus domestica* (apple tree)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C;Accession: T17014
R;Reid, S.J.; Ross, G.S.
Physiol. Plantarum 100, 183-189, 1997
A;Title: Up-regulation of two cDNA clones encoding metallothionein-like proteins in apple fruit during cool storage.
A;Reference number: Z18652
A;Accession: T17014
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-79 <REI>
A;Cross-references: EMBL:U61973; NID:g1655850; PID:g1655851
A;Experimental source: apple flesh cortical tissue
C;Genetics:
A;Gene: AMT1

C;Superfamily: metallothionein
C;Keywords: metal binding

Query Match 10.6%; Score 45; DB 2; Length 79;
Best Local Similarity 30.6%; Pred. No. 4.3e+02;
Matches 15; Conservative 7; Mismatches 21; Indels 6; Gaps 2;

Qy 4 SGCSSQSISPMRSISENS--LVAMDFSGQKSRVIENPTEALSVAVEEG 49
|||: :| | | : | : ||| : : || | |
Db 19 SGCNGCGMAPDLSYMEGSTTETLVMGVAPQKSHM---EASEMGVAAENG 64

RESULT 13

D69087

hydrogenase expression/formation protein HypC - Methanobacterium
thermoautotrophicum (strain Delta H)

C;Species: Methanobacterium thermoautotrophicum

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 14-Apr-2003

C;Accession: D69087

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.;
Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.;
Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire,
R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.;
Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.;
Petrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.;
Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta
H: functional analysis and comparative genomics.

A;Reference number: A69000; MUID:98037514; PMID:9371463

A;Accession: D69087

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-82 <MTH>

A;Cross-references: GB:AE000924; GB:AE000666; NID:g2622777; PIDN:AAB86122.1;
PID:g2622778

A;Experimental source: strain Delta H

C;Genetics:

A;Gene: MTH1649

C;Superfamily: [NiFe]-hydrogenase maturation chaperone

Query Match 10.6%; Score 45; DB 2; Length 82;
Best Local Similarity 28.9%; Pred. No. 4.5e+02;
Matches 11; Conservative 9; Mismatches 14; Indels 4; Gaps 1;

Qy 18 SENSLVAMDFSGQKSRV----IENPTEALSVAVEEGLA 51
||::: :|| | : :| ::: | | | | |
Db 14 SEDNIATVDFGVRQQVKLDLVDVVEEGKYVLVHSGYA 51

RESULT 14

AB2271

periplasmic mercuric ion binding protein [imported] - Nostoc sp. (strain PCC
7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AB2271
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;
 Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.;
 Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.;
 Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing
 Cyanobacterium Anabaena sp. strain PCC 7120.
 A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Accession: AB2271
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-64 <KUR>
 A;Cross-references: GB:BA000019; PIDN:BAB75420.1; PID:g17132855; GSPDB:GN00179
 A;Experimental source: strain PCC 7120
 C;Genetics:
 A;Gene: asl3721

Query Match 10.5%; Score 44.5; DB 2; Length 64;
 Best Local Similarity 16.7%; Pred. No. 3.8e+02;
 Matches 8; Conservative 20; Mismatches 17; Indels 3; Gaps 1;

Qy 4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLA 51
 | |:: : :::: ::|| | :::: |::| :::: ||
 Db 13 SACANNITNAVKTVDVAIVQAD---PQTKLVNVETQASETSIKDALA 57

RESULT 15
 S75293
 hypothetical protein ssr2333 - Synechocystis sp. (strain PCC 6803)
 C;Species: Synechocystis sp.
 A;Variety: PCC 6803
 C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C;Accession: S75293
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;
 Miyajima, N.; Hirose, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.;
 Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.;
 Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and
 assignment of potential protein-coding regions.
 A;Reference number: S74322; MUID:97061201; PMID:8905231
 A;Accession: S75293
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-79 <KAN>
 A;Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BAA17207.1;
 PID:d1017940; PID:g1652284
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June
 1996

Query Match 10.5%; Score 44.5; DB 2; Length 79;
 Best Local Similarity 22.1%; Pred. No. 4.9e+02;
 Matches 15; Conservative 14; Mismatches 34; Indels 5; Gaps 1;

Qy 16 SISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGLRLGTHGSPTASSQSSAT 75

```

      | : : : |      | : : | || : : : : | |      | |      | |
Db      7 SVGQLAFVEKILLGNHGQGLVNRLEAMGIIPDKPIQLLRKAGL-----GGPLHLRIGSTT 61

QY      76 NMAIHRSQ 83
      :| : ||:
Db      62 EVAMRRSE 69

```

RESULT 16

D85807

hypothetical protein Z2988 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C;Accession: D85807

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.
Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: D85807

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-51 <STO>

A;Cross-references: GB:AE005174; NID:g12516000; PIDN:AAG56920.1; GSPDB:GN00145; UWGP:Z2988

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: Z2988

```

Query Match          10.4%;  Score 44;  DB 2;  Length 51;
Best Local Similarity 32.0%;  Pred. No. 3.3e+02;
Matches      8;  Conservative      6;  Mismatches    11;  Indels      0;  Gaps      0;

```

```

QY      37 NPTEALSVAVEEGLAWRKKGCLRLG 61
      :| | : | :||:: || |

```

```

Db      20 SPAEIFMMTPGEVVSWRERAAALRSG 44

```

RESULT 17

B90959

probable phage tail protein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C;Accession: B90959

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida, T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: B90959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-78 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB36065.1; PID:g13362110; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2642

Query Match 10.4%; Score 44; DB 2; Length 78;
Best Local Similarity 32.0%; Pred. No. 5.5e+02;
Matches 8; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Qy 37 NPTEALSVAVEEGLAWRKKGCLRLG 61
:| | : | :||: || |
Db 47 SPAEIFMMTPGEVVSWRERAALRSG 71

RESULT 18

A75411

hypothetical protein - *Deinococcus radiodurans* (strain R1)

C;Species: *Deinococcus radiodurans*

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C;Accession: A75411

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.;
Dodson, R.J.; Haft, D.H.; Gwinn, M.L.; Nelson, W.C.; Richardson, D.L.; Moffat,
K.S.; Qin, H.; Jiang, L.; Pamphile, W.; Crosby, M.; Shen, M.; Vamathevan, J.J.;
Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Makarova, K.S.; Aravind, L.;
Daly, M.J.; Minton, K.W.; Fleischmann, R.D.; Ketchum, K.A.; Nelson, K.E.;
Salzberg, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: A75411

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-78 <WHI>

A;Cross-references: GB:AE001978; GB:AE000513; NID:g6459059; PIDN:AAF10892.1;
PID:g6459071; TIGR:DR1317; GSPDB:GN00077

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR1317

A;Map position: 1

Query Match 10.4%; Score 44; DB 2; Length 78;
Best Local Similarity 21.4%; Pred. No. 5.5e+02;
Matches 12; Conservative 13; Mismatches 23; Indels 8; Gaps 1;

Qy 21 SLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATN 76
|| | | | : | | : : : || : : | : : : : | :
Db 10 SLAAFDNGAMKKAVLAVPALLLALSL-----SGCQKQADSNTSTSTTTTKSTD 57

RESULT 19

G90914

A;Map position: REV504744-504496

C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0567

Query Match 10.2%; Score 43; DB 1; Length 82;
Best Local Similarity 25.9%; Pred. No. 7.6e+02;
Matches 15; Conservative 10; Mismatches 15; Indels 18; Gaps 3;

Qy 4 SGCSSQSISPMRSISENSLVAMDFS-GQKSRVIEN-----PTEALSVAVEEGLAWR 53
:| | : ||:| : | | :| | | | : :| | : || | :
Db 20 AGCGAM-----QRLVSMGINIGSKLKVIRNQNGPVIISTKGSNIAIGRGLAMK 67

RESULT 21

D82682

hypothetical protein XF1429 [imported] - Xylella fastidiosa (strain 9a5c)

C;Species: Xylella fastidiosa

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C;Accession: D82682

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: D82682

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-52 <SIM>

A;Cross-references: GB:AE003973; GB:AE003849; NID:g9106438; PIDN:AAF84238.1;

GSPDB:GN00128; XFSC:XF1429

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros, M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van

hypothetical protein 56 - Staphylococcus aureus phage phi PVL
 C;Species: Staphylococcus aureus phage phi PVL
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C;Accession: T00189
 R;Kaneko, J.; Kimura, T.; Kawakami, Y.; Tomita, T.; Kamio, Y.
 Biosci. Biotechnol. Biochem. 61, 1960-1962, 1997
 A;Title: Panton-Valentine leukocidin genes in a phage-like particle isolated
 from mitomycin C-treated Staphylococcus aureus V8(ATCC 49775).
 A;Reference number: Z14119; MUID:98067870; PMID:9404084
 A;Accession: T00189
 A;Status: translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-68 <KAN>
 A;Cross-references: EMBL:AB009866; NID:d1204727; PIDN:BAA31929.1; PID:d1032890

Query Match 10.0%; Score 42.5; DB 2; Length 68;
 Best Local Similarity 29.2%; Pred. No. 6.9e+02;
 Matches 14; Conservative 11; Mismatches 20; Indels 3; Gaps 2;

Qy 7 SSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVE--EGLAW 52
 :| || : :: | | :| || :|| |:: :| |
 Db 20 ASNEISELLYEYDSELMSADEDGD-NRDIEEKRDALKQAIQIIDKLTW 66

RESULT 24

S72807

hypothetical protein B1549_F3_145 - Mycobacterium leprae
 C;Species: Mycobacterium leprae
 C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
 C;Accession: S72807
 R;Smith, D.R.; Robison, K.
 submitted to the EMBL Data Library, November 1993
 A;Description: Mycobacterium leprae cosmid B1549.
 A;Reference number: S72582
 A;Accession: S72807
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-74 <SMI>
 A;Cross-references: EMBL:U00014; NID:g466903; PIDN:AAA50904.1; PID:g466929
 C;Genetics:
 A;Start codon: GTG

Query Match 10.0%; Score 42.5; DB 2; Length 74;
 Best Local Similarity 31.4%; Pred. No. 7.6e+02;
 Matches 11; Conservative 6; Mismatches 13; Indels 5; Gaps 1;

Qy 38 PTEALSVAVE-----EGLAWRKKGCLRLGTHGSPT 67
 | | ::||: | || : | ||:| |
 Db 22 PVLATAIALQLTSENECAQWRLGETVTLHEHGNPT 56

RESULT 25

JC2006

differentiation inhibitor Id2B - human
 C;Species: Homo sapiens (man)
 C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 29-Sep-1999
 C;Accession: JC2006

R;Kurabayashi, M.; Jeyaseelan, R.; Kedes, L.

Gene 133, 305-306, 1993

A;Title: Two distinct cDNA sequences encoding the human helix-loop-helix protein Id2.

A;Reference number: JC2006; MUID:94040830; PMID:8224921

A;Accession: JC2006

A;Molecule type: mRNA

A;Residues: 1-36 <KUR>

A;Cross-references: GB:M96843; NID:g397775; PIDN:AAA16865.1; PID:g397776

A;Experimental source: heart

C;Superfamily: transcription repressor Id-2

Query Match 9.9%; Score 42; DB 2; Length 36;

Best Local Similarity 53.3%; Pred. No. 3.7e+02;

Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 9 QSISPMRSISENSLV 23

:: ||:|||| :|||:

Db 2 KAFSPVRSIRKNSLL 16

Search completed: July 8, 2004, 08:20:47

Job time : 19.8583 secs

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:20:54 ; Search time 54.2362 Seconds
(without alignments)
483.093 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSI SPMR SISEN.....SPTASSQSSATNMAIHR SQP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1279676 seqs, 311918243 residues

Total number of hits satisfying chosen parameters: 487241

Minimum DB seq length: 0
Maximum DB seq length: 85

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score Match Length DB ID	Description

1	64	15.1	75	12	US-10-424-599-234105	Sequence 234105,
2	56.5	13.4	78	15	US-10-121-016-49	Sequence 49, Appl
3	56	13.2	79	11	US-09-833-245-1961	Sequence 1961, Ap
4	55	13.0	67	12	US-10-424-599-212021	Sequence 212021,
5	54.5	12.9	53	12	US-10-276-774-1962	Sequence 1962, Ap
6	54.5	12.9	60	9	US-09-925-299-1024	Sequence 1024, Ap
7	54.5	12.9	60	10	US-09-925-299-1024	Sequence 1024, Ap
8	54.5	12.9	77	16	US-10-437-963-148035	Sequence 148035,
9	53	12.5	69	14	US-10-029-386-31268	Sequence 31268, A
10	52.5	12.4	65	16	US-10-437-963-166254	Sequence 166254,
11	52	12.3	53	15	US-10-074-024-278	Sequence 278, App
12	52	12.3	65	12	US-10-424-599-199787	Sequence 199787,
13	52	12.3	78	12	US-10-424-599-229841	Sequence 229841,
14	51.5	12.2	81	12	US-10-424-599-219466	Sequence 219466,
15	51.5	12.2	83	9	US-09-908-711-103	Sequence 103, App
16	51.5	12.2	83	10	US-09-764-891-3234	Sequence 3234, Ap
17	51	12.1	72	10	US-09-764-891-4413	Sequence 4413, Ap
18	51	12.1	72	16	US-10-437-963-138148	Sequence 138148,
19	51	12.1	84	9	US-09-864-761-47521	Sequence 47521, A
20	51	12.1	84	16	US-10-437-963-141837	Sequence 141837,
21	50.5	11.9	55	16	US-10-437-963-117556	Sequence 117556,
22	50.5	11.9	80	14	US-10-125-258-6	Sequence 6, Appli
23	50.5	11.9	83	15	US-10-264-049-3394	Sequence 3394, Ap
24	50	11.8	75	16	US-10-437-963-125147	Sequence 125147,
25	50	11.8	77	12	US-10-424-599-149927	Sequence 149927,
26	50	11.8	84	16	US-10-437-963-197341	Sequence 197341,
27	49.5	11.7	47	12	US-10-424-599-210837	Sequence 210837,
28	49.5	11.7	60	12	US-10-424-599-168279	Sequence 168279,
29	49.5	11.7	80	11	US-09-833-245-1403	Sequence 1403, Ap
30	49.5	11.7	80	11	US-09-833-245-1405	Sequence 1405, Ap
31	49.5	11.7	80	16	US-10-437-963-202693	Sequence 202693,
32	49	11.6	36	12	US-10-424-599-174085	Sequence 174085,
33	49	11.6	66	16	US-10-437-963-160296	Sequence 160296,
34	49	11.6	75	16	US-10-437-963-103570	Sequence 103570,
35	48.5	11.5	51	10	US-09-764-891-5151	Sequence 5151, Ap
36	48.5	11.5	65	12	US-10-424-599-258371	Sequence 258371,
37	48.5	11.5	73	14	US-10-106-698-6971	Sequence 6971, Ap
38	48.5	11.5	76	11	US-09-864-408A-6504	Sequence 6504, Ap
39	48.5	11.5	77	12	US-10-424-599-221632	Sequence 221632,
40	48.5	11.5	81	16	US-10-437-963-103261	Sequence 103261,
41	48.5	11.5	82	9	US-09-864-761-42490	Sequence 42490, A
42	48	11.3	64	16	US-10-437-963-182656	Sequence 182656,
43	48	11.3	73	12	US-10-424-599-156796	Sequence 156796,
44	48	11.3	76	12	US-10-282-122A-65522	Sequence 65522, A
45	48	11.3	76	16	US-10-437-963-131807	Sequence 131807,
46	48	11.3	84	16	US-10-437-963-119477	Sequence 119477,
47	48	11.3	85	12	US-10-424-599-178208	Sequence 178208,
48	47.5	11.2	63	14	US-10-029-386-29148	Sequence 29148, A
49	47.5	11.2	77	12	US-10-221-278-278	Sequence 278, App
50	47.5	11.2	77	15	US-10-291-172-278	Sequence 278, App
51	47	11.1	45	16	US-10-437-963-128337	Sequence 128337,
52	47	11.1	46	12	US-10-424-599-225061	Sequence 225061,
53	47	11.1	60	16	US-10-437-963-175457	Sequence 175457,
54	47	11.1	64	15	US-10-074-024-236	Sequence 236, App
55	47	11.1	68	16	US-10-437-963-202744	Sequence 202744,
56	47	11.1	78	16	US-10-437-963-164448	Sequence 164448,

57	47	11.1	85	12	US-10-424-599-182217	Sequence 182217,
58	46.5	11.0	69	12	US-10-424-599-177050	Sequence 177050,
59	46.5	11.0	78	9	US-09-764-887-218	Sequence 218, App
60	46.5	11.0	78	14	US-10-073-961-218	Sequence 218, App
61	46.5	11.0	78	16	US-10-437-963-102594	Sequence 102594,
62	46.5	11.0	79	12	US-10-424-599-191756	Sequence 191756,
63	46.5	11.0	82	16	US-10-437-963-178917	Sequence 178917,
64	46.5	11.0	83	16	US-10-437-963-158251	Sequence 158251,
65	46	10.9	52	11	US-09-864-408A-7624	Sequence 7624, Ap
66	46	10.9	62	12	US-10-424-599-216588	Sequence 216588,
67	46	10.9	64	12	US-10-424-599-150131	Sequence 150131,
68	46	10.9	68	9	US-09-764-878-131	Sequence 131, App
69	46	10.9	68	9	US-09-764-860-468	Sequence 468, App
70	46	10.9	68	14	US-10-079-854-131	Sequence 131, App
71	46	10.9	68	14	US-10-074-095-468	Sequence 468, App
72	46	10.9	68	15	US-10-212-872-468	Sequence 468, App
73	46	10.9	69	9	US-09-764-847-700	Sequence 700, App
74	46	10.9	69	14	US-10-092-154-700	Sequence 700, App
75	46	10.9	70	14	US-10-268-518-5	Sequence 5, Appli
76	46	10.9	71	12	US-10-424-599-255309	Sequence 255309,
77	46	10.9	75	16	US-10-437-963-190577	Sequence 190577,
78	46	10.9	77	16	US-10-437-963-191204	Sequence 191204,
79	46	10.9	81	12	US-10-424-599-196302	Sequence 196302,
80	46	10.9	85	16	US-10-437-963-126461	Sequence 126461,
81	45.5	10.8	50	16	US-10-437-963-125673	Sequence 125673,
82	45.5	10.8	53	12	US-10-424-599-271608	Sequence 271608,
83	45.5	10.8	55	16	US-10-437-963-162380	Sequence 162380,
84	45.5	10.8	56	9	US-09-925-299-1460	Sequence 1460, Ap
85	45.5	10.8	56	10	US-09-925-299-1460	Sequence 1460, Ap
86	45.5	10.8	58	9	US-09-925-301-1647	Sequence 1647, Ap
87	45.5	10.8	58	9	US-09-925-299-1372	Sequence 1372, Ap
88	45.5	10.8	58	9	US-09-925-299-1437	Sequence 1437, Ap
89	45.5	10.8	58	9	US-09-925-299-1493	Sequence 1493, Ap
90	45.5	10.8	58	9	US-09-925-299-1528	Sequence 1528, Ap
91	45.5	10.8	58	10	US-09-925-299-1372	Sequence 1372, Ap
92	45.5	10.8	58	10	US-09-925-299-1437	Sequence 1437, Ap
93	45.5	10.8	58	10	US-09-925-299-1493	Sequence 1493, Ap
94	45.5	10.8	58	10	US-09-925-299-1528	Sequence 1528, Ap
95	45.5	10.8	59	16	US-10-437-963-203339	Sequence 203339,
96	45.5	10.8	60	9	US-09-764-877-1350	Sequence 1350, Ap
97	45.5	10.8	60	15	US-10-242-515-1350	Sequence 1350, Ap
98	45.5	10.8	66	12	US-10-424-599-266852	Sequence 266852,
99	45.5	10.8	71	12	US-10-424-599-240469	Sequence 240469,
100	45.5	10.8	72	9	US-09-864-761-38891	Sequence 38891, A

ALIGNMENTS

RESULT 1

US-10-424-599-234105
 ; Sequence 234105, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234105
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(75)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53420C.1.pep
US-10-424-599-234105

Query Match 15.1%; Score 64; DB 12; Length 75;
Best Local Similarity 37.5%; Pred. No. 5.4;
Matches 21; Conservative 7; Mismatches 22; Indels 6; Gaps 3;

Qy 34 VIENPTEALSVAVEEGLAWRKKGCLR--LGTHGSPTASSQSSA---TNMAIHRSQP 84
||| : || : ||| : | : | : ||| : |
Db 5 VIHN-XKHCEVAKKRILFWRKRCVNGPTGRNERTDPSAQSSAEYLTXPAIHKGNP 59

RESULT 2

US-10-121-016-49

; Sequence 49, Application US/10121016
; Publication No. US20040010811A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Pia M. Challita-Eid
; APPLICANT: Arthur B. Raitano
; APPLICANT: Mary Faris
; APPLICANT: Rene S. Hubert
; APPLICANT: Karen Jane Meyrick Morrison
; APPLICANT: Robert Kendall Morrison
; APPLICANT: Wangmao Ge
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
; TITLE OF INVENTION: ENTITLED 162P1E6 USEFUL IN TREATMENT AND DETECTION OF
CANCER
; FILE REFERENCE: 51158-20077.00
; CURRENT APPLICATION NUMBER: US/10/121,016
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/283,112
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/286,630
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 78

; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-121-016-49

Query Match 13.4%; Score 56.5; DB 15; Length 78;
Best Local Similarity 34.8%; Pred. No. 47;
Matches 16; Conservative 6; Mismatches 23; Indels 1; Gaps 1;

Qy 40 EALSVAVEEGLAWRKKGCLR-LGTHGSPTASSQSSATNMAIHRSQP 84
|| | | : || ||: | ::|||: ||:||
Db 33 EAADGHPEMGFHHATQACLELLGSSDLPASASQSAGITGVNHRAQP 78

RESULT 3

US-09-833-245-1961
; Sequence 1961, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1961
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245--1961

Query Match 13.2%; Score 56; DB 11; Length 79;
Best Local Similarity 39.3%; Pred. No. 55;
Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 57 CLRLGTHGSPTASSQSSATNMAIHRSQP 84
|| :| | |: | | : | |||
Db 45 CLSIGQHELPSYSCQPGRKRLLPHHSQP 72

RESULT 4

US-10-424-599-212021
; Sequence 212021, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 212021
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_33481C.1.pep
US-10-424-599-212021

Query Match 13.0%; Score 55; DB 12; Length 67;
Best Local Similarity 29.3%; Pred. No. 59;
Matches 22; Conservative 7; Mismatches 26; Indels 20; Gaps 2;

Qy 4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTH 63
:| || | :|: | || | || || |
Db 9 TGFSSVHFSSRAYKQASMLLALFAGGDGYRVEEN-----NGCLMLGWH 51

Qy 64 GSP---TASSQSSAT 75
| |: | :||
Db 52 TRPLIATSAWQLAAT 66

RESULT 5

US-10-276-774-1962

; Sequence 1962, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1962
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-1962

Query Match 12.9%; Score 54.5; DB 12; Length 53;
Best Local Similarity 36.7%; Pred. No. 51;
Matches 18; Conservative 9; Mismatches 17; Indels 5; Gaps 3;

Qy 39 TEALSVAVEEGLAWRKKGCLR--LGTHGSP-TASSQSSATNMAIHRSQ 83
||: ||| : |: || |: |: ||| :|| : | ||::
Db 5 TESRSVA-QAGVQWRDLSSLQPPPPGSRGSPASASPVAGITGTRHHRTR 52

RESULT 6

US-09-925-299-1024

```

; Sequence 1024, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1024
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (26)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (38)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1024

```

```

Query Match      12.9%; Score 54.5; DB 9; Length 60;
Best Local Similarity 50.0%; Pred. No. 59;
Matches 14; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

```

```

Qy      58 LRLGTHGSPTAS-SQSSATNMAIHRSQP 84
        | | | | | | | | | | | | | |
Db      33 LELATXGDPPASASQSGGITGVSHRAQP 60

```

RESULT 7

US-09-925-299-1024

```

; Sequence 1024, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299

```

```

; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1024
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (26)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (38)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1024

```

```

Query Match          12.9%; Score 54.5; DB 10; Length 60;
Best Local Similarity 50.0%; Pred. No. 59;
Matches 14; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

```

```

Qy      58 LRLGTHGSPTAS-SQSSATNMAIHRSP 84
        | | | | | | | | | | | | | |
Db      33 LELATXGDPPASASQSGGITGVSHRAQP 60

```

```

RESULT 8
US-10-437-963-148035
; Sequence 148035, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14

```


US-10-437-963-166254

US-10-437-963-166254

Matches 16; Conservative 12; Mismatches 31; Indels 5; Gaps 1;

Ov 76 NMAI 79

US-10-074-024-278

; LENGTH: 53

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (24)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-074-024-278

Query Match 12.3%; Score 52; DB 15; Length 53;
Best Local Similarity 43.5%; Pred. No. 1e+02;
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 59 RLGTHGSPTASSQSSATNMAIHR 81
|: || :| || | |:|
Db 18 RVAVHGXSPPTSVSSLTERAVHR 40

RESULT 12

US-10-424-599-199787

; Sequence 199787, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 199787

; LENGTH: 65

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_22431C.1.pep

US-10-424-599-199787

Query Match 12.3%; Score 52; DB 12; Length 65;
Best Local Similarity 27.3%; Pred. No. 1.3e+02;
Matches 18; Conservative 12; Mismatches 28; Indels 8; Gaps 3;

Qy 1 QGRSGCSSQSI SPMRIS ENSLVAMDFSGQKSRVIENPTE---ALSVAVEEGLAWRKKGC 57
||| | | | | | :|:: : : : | : || : : : | : |
Db 5 QGRRACYSLSTSKMS--GEEGVIAVEPAAAAAIPGEPMDIMTALQLVLRKSLGY---GW 59

Qy 58 LRLGTH 63

| | |

Db 60 LSRGLH 65

```

US-10-424-599-229841
; Sequence 229841, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 229841
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_49570C.1.pep
US-10-424-599-229841

```

```

Qy      23 VAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCRLGLTHGSPTASSQSS 73
      : : | : | : | : | | : | | | | | : | | : : |
Db      9  ILLDKTGDWITLIQNSTISLGSRV----AARKKGCGLTVMIVMSGNAAGKRS 55

```

```

US-10-424-599-219466
; Sequence 219466, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 219466
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_40203C.1.pep
US-10-424-599-219466

```


Query Match 12.2%; Score 51.5; DB 12; Length 81;
Best Local Similarity 23.9%; Pred. No. 2e+02;
Matches 16; Conservative 11; Mismatches 31; Indels 9; Gaps 2;

Qy 22 LVAMDFSGQKSRVIENPTEALSVAVEEG----LAWRKKGCLRLGTHGSPTASSQSSATNM 77
|::||| : : || || || | : | :
Db 1 LLSHDHSAYKLEHVTHENRNEQERVREGNADRLGWLKNGC-----HPNGFLKKQRAGMHF 55

Qy 78 AIHRSQP 84
:|: :|
Db 56 SINNTKP 62

RESULT 15

US-09-908-711-103

; Sequence 103, Application US/09908711

; Patent No. US20020045230A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA128

; CURRENT APPLICATION NUMBER: US/09/908,711

; CURRENT FILING DATE: 2001-07-20

; PRIOR APPLICATION NUMBER: US01/01360

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 09/764,867

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01344

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 09/764,892

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01345

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 09/764,888

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01329

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 09/764,905

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01354

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 09/764,891

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01339

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 09/764,869

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01340

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 09/764,874

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01334

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 09/764,898

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01320

; PRIOR FILING DATE: 2001-01-17

```

; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
;   LENGTH: 83
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (25)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-103

```

```

Query Match          12.2%; Score 51.5; DB 9; Length 83;
Best Local Similarity 27.5%; Pred. No. 2.1e+02;
Matches    19; Conservative    8; Mismatches    33; Indels    9; Gaps    2;

Qy      15 RSISENSLVAMDF-SGQKSRVIENPTEALSVAVEEGLAWRKK-----GCLRLGTHGS 65
      | : | | | : || || | : ||:: | | | :
Db      9 RDVGEGLDLPQMEVGSGXGSRPRTPPASGPRHSSRRKAPWRRRLPSQWWNPGGTRPGSAAQ 68

```

Qy 66 PTASSQSSA 74
 | |||:|:
 Db 69 PWGSSQASS 77

RESULT 16

US-09-764-891-3234

; Sequence 3234, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3234
 ; LENGTH: 83
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (25)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-891-3234

Query Match 12.2%; Score 51.5; DB 10; Length 83;
 Best Local Similarity 27.5%; Pred. No. 2.1e+02;
 Matches 19; Conservative 8; Mismatches 33; Indels 9; Gaps 2;

Qy 15 RSISENSLVAMDF-SGQKSRVIENPTEALSVAVEEGLAWRKK-----GCLRLGTHGS 65
 | : | | |: || || | : ||:: | | |:
 Db 9 RDVGEGLDPQMEVGSGXGSRPRTPPASGPRHSSRRKAPWRRRLPSQWWNPGGTRPGSAAQ 68

Qy 66 PTASSQSSA 74
 | |||:|:
 Db 69 PWGSSQASS 77

RESULT 17

US-09-764-891-4413

; Sequence 4413, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4413
 ; LENGTH: 72

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (21)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4413

Query Match 12.1%; Score 51; DB 10; Length 72;
Best Local Similarity 35.6%; Pred. No. 2e+02;
Matches 16; Conservative 8; Mismatches 17; Indels 4; Gaps 2;

Qy 39 TEALSVAVEEGLAWRKKGCLR---LGTHGSPTASSQSSATNMAIH 80
|| ||: : |: | | |: ||: || ::| : | ||
Db 19 TEXCSVS-QAGVQWPDFGSLQLRLLG SCHSPASASGVAGTTGACH 62

RESULT 18

US-10-437-963-138148

; Sequence 138148, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules

Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 138148

; LENGTH: 72

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_39563C.1.pep

US-10-437-963-138148

Query Match 12.1%; Score 51; DB 16; Length 72;
Best Local Similarity 22.1%; Pred. No. 2e+02;
Matches 19; Conservative 14; Mismatches 23; Indels 30; Gaps 3;

Qy 7 SSQSI SPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSP 66
| | || |: :|| : || :||: |:: | | | |
Db 4 SHDS DSPAASLPASSL-----LRRPTYSLALV-----REQPSLATGAHQQP 44

Qy 67 TASS-----QSSATNMAIHR 81
: | :| ::::||
Db 45 RSQSLPKASPRSLALVGTALSLSVHR 70

RESULT 19

US-09-864-761-47521

; Sequence 47521, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES

USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 47521

; LENGTH: 84

; TYPE: PRT

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL158153.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.7
; OTHER INFORMATION: EST_HUMAN HIT: BF573955.1, EVALUE 1.60e-02
; OTHER INFORMATION: SWISSPROT HIT: Q91641, EVALUE 3.00e-25
US-09-864-761-47521

Query Match 12.1%; Score 51; DB 9; Length 84;
Best Local Similarity 27.8%; Pred. No. 2.5e+02;
Matches 15; Conservative 6; Mismatches 7; Indels 26; Gaps 2;

Qy 29 GQKSRVIENP-----TEALSVAV-----EEGLAWRKKG 56
|||:|:: | | ||: || | ||::|
Db 11 GQKARLLSRPLRGVSGKHCLTFFYHMYGGGTGLLSVYLKKEEDSEESLLWRRRG 64

RESULT 20

US-10-437-963-141837
; Sequence 141837, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 141837
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(84)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_42902C.1.pep
US-10-437-963-141837

Query Match 12.1%; Score 51; DB 16; Length 84;
Best Local Similarity 35.8%; Pred. No. 2.5e+02;
Matches 19; Conservative 6; Mismatches 24; Indels 4; Gaps 2;

Qy 30 QKSRVIENPTEALSVAV--EEGLAWRKKGCLRLGTHGSPTAS--SQSSATNMA 78
:| | || | | :|: ||:| | | :|||:|
Db 18 KKRRGDWNPCFRLQVLTCKHRPLSSRRAPCLKLLAHGQRIVSLADIASATNLA 70

RESULT 21

US-10-437-963-117556
 ; Sequence 117556, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
 Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 117556
 ; LENGTH: 55
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_20950C.1.pep
 US-10-437-963-117556

Query Match 11.9%; Score 50.5; DB 16; Length 55;
 Best Local Similarity 42.1%; Pred. No. 1.6e+02;
 Matches 16; Conservative 4; Mismatches 15; Indels 3; Gaps 2;

Qy 39 TEALSVAVEEGLAWRKKGCLR-LGTHGSPTASSQSSAT 75
 | :: || ||| | || | | |:: |||
 Db 2 TSTMTAAV--GLAWSGAGWLRGAGAAGLTAATTGRSAT 37

RESULT 22

US-10-125-258-6
 ; Sequence 6, Application US/10125258
 ; Publication No. US20030028920A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Altier, Daniel J.
 ; APPLICANT: Herrmann, Rafael
 ; APPLICANT: Lu, Albert L.
 ; APPLICANT: McCutchen, Billy F.
 ; APPLICANT: Presnail, James K.
 ; APPLICANT: Weaver, Janine L.
 ; APPLICANT: Wong, James F. H.
 ; TITLE OF INVENTION: Antimicrobial Polypeptides and Their
 ; TITLE OF INVENTION: Uses
 ; FILE REFERENCE: 35718/246215
 ; CURRENT APPLICATION NUMBER: US/10/125,258
 ; CURRENT FILING DATE: 2002-04-18
 ; PRIOR APPLICATION NUMBER: 60/285,355

; PRIOR FILING DATE: 2001-04-20
 ; NUMBER OF SEQ ID NOS: 127
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 80
 ; TYPE: PRT
 ; ORGANISM: Manduca sexta
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: 78
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-10-125-258-6

Query Match 11.9%; Score 50.5; DB 14; Length 80;
 Best Local Similarity 28.8%; Pred. No. 2.7e+02;
 Matches 17; Conservative 10; Mismatches 23; Indels 9; Gaps 1;

Qy 16 SISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSA 74
 |:| |||: |:| | : : : |:| |||| |:| ::
 Db 2 SLSCLFLVALALVGAESRYIADDVVLVPMVSR-----VRRDTHGSVTVNSDGTS 51

RESULT 23

US-10-264-049-3394
 ; Sequence 3394, Application US/10264049
 ; Publication No. US20040005579A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PA133P1
 ; CURRENT APPLICATION NUMBER: US/10/264,049
 ; CURRENT FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/18569
 ; PRIOR FILING DATE: 2001-06-07
 ; PRIOR APPLICATION NUMBER: US 60/209,467
 ; PRIOR FILING DATE: 2000-06-07
 ; NUMBER OF SEQ ID NOS: 4360
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 3394
 ; LENGTH: 83
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-264-049-3394

Query Match 11.9%; Score 50.5; DB 15; Length 83;
 Best Local Similarity 30.0%; Pred. No. 2.8e+02;
 Matches 18; Conservative 7; Mismatches 18; Indels 17; Gaps 3;

Qy 35 IENPTEALSVAVEEGL-----AWRKKGCLRLGTHGS-----PTASSQSSATNMAIH 80
 :: |||| | :|| || : | | | | | | | | : | :|:
 Db 13 VQKPTEAQS---RQGLTDLCWYLGAWIAELSLLEGKWGGVGGPDRPGCQSASAKTTLAVH 69

RESULT 24

US-10-437-963-125147
 ; Sequence 125147, Application US/10437963
 ; Publication No. US20040123343A1


```
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 125147
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_2781C.1.pep
US-10-437-963-125147
```

```
Query Match          11.8%; Score 50; DB 16; Length 75;
Best Local Similarity 35.7%; Pred. No. 2.8e+02;
Matches 15; Conservative 5; Mismatches 14; Indels 8; Gaps 2;
```

```
Qy      39 TEALSVAVEEGLAW-----RKKGCLRLGTHGSPTASSQSSAT 75
      |||  ||  |      |: | ||| | || :: : |
Db      12 TEA---GVESDDPWPARRTRRGVCARLGRGGFPTRTARGTVT 50
```

```
RESULT 25
US-10-424-599-149927
; Sequence 149927, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 149927
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_106405C.1.pep
US-10-424-599-149927
```

Query Match 11.8%; Score 50; DB 12; Length 77;
Best Local Similarity 25.9%; Pred. No. 2.9e+02;
Matches 14; Conservative 4; Mismatches 22; Indels 14; Gaps 1;

```
Qy      40 EALSVAVEEGLAW-----RKKGCLRLGTHGSPTASSQSSATNMAI 79
      | || |: : |           ||| |  || :  || |:
Db      18 EGLSCAINKSTTWYKVDTYIIFLDKIRRKKRALMFSGHIAASKRNSSIMKRAL 71
```

Search completed: July 8, 2004, 08:31:42
Job time : 56.2362 secs

Result						Query	
No.	Score	Match	Length	DB	ID	Description	

1	62.5	14.8	67	16	Q93J23	Q93j23 streptomyce
2	57	13.5	78	16	Q7VBG2	Q7vbg2 prochloroco
3	56	13.2	58	9	O80316	O80316 bacterioph
4	55	13.0	75	5	Q8IRX0	Q8irx0 drosophila
5	54	12.8	73	16	Q89Y84	Q89y84 bradyrhizob
6	54	12.8	80	5	Q23341	Q23341 caenorhabdi
7	54	12.8	84	16	Q9JMT2	Q9jmt2 escherichia
8	53	12.5	80	12	Q91LH8	Q91lh8 white spot
9	51	12.1	82	9	Q854J6	Q854j6 mycobacteri
10	50	11.8	72	9	Q853P2	Q853p2 mycobacteri
11	49	11.6	73	16	Q7VCF8	Q7vcf8 prochloroco
12	48	11.3	61	10	Q8H431	Q8h431 oryza sativ
13	48	11.3	64	10	Q8LRC4	Q8lrc4 oryza sativ
14	47.5	11.2	66	15	Q9J125	Q9j125 human immun
15	47	11.1	64	5	Q8MMG0	Q8mmg0 drosophila
16	47	11.1	68	16	Q8E7B7	Q8e7b7 streptococc
17	47	11.1	72	10	Q84Z53	Q84z53 oryza sativ
18	47	11.1	79	11	Q8R0X3	Q8r0x3 mus musculu
19	46.5	11.0	73	16	Q8EF85	Q8ef85 shewanella
20	46.5	11.0	79	2	Q9RCD4	Q9rcd4 xanthomonas
21	46.5	11.0	80	3	Q96U90	Q96u90 neurospora
22	46.5	11.0	80	16	Q7V3F4	Q7v3f4 prochloroco
23	46.5	11.0	83	12	Q8B6L1	Q8b6l1 soybean dwa
24	46.5	11.0	83	12	Q8B6K9	Q8b6k9 soybean dwa
25	46.5	11.0	83	16	Q8EIV8	Q8eiv8 shewanella
26	46	10.9	54	10	Q8VY75	Q8vy75 arabidopsis
27	46	10.9	77	15	Q76048	Q76048 human immun
28	45.5	10.8	65	16	Q9PCX0	Q9pcx0 xylella fas
29	45.5	10.8	72	16	Q8NLI4	Q8nli4 corynebacte
30	45.5	10.8	73	5	P91302	P91302 caenorhabdi
31	45.5	10.8	79	16	Q8EW35	Q8ew35 mycoplasma
32	45.5	10.8	81	2	P94624	P94624 clostridium
33	45	10.6	59	9	Q855Q7	Q855q7 mycobacteri
34	45	10.6	60	12	Q89224	Q89224 vaccinia vi
35	45	10.6	69	15	Q9J151	Q9j151 human immun
36	45	10.6	69	16	Q7UEB9	Q7ueb9 rhodopirell
37	45	10.6	73	13	Q8JHU0	Q8jhu0 gallus gall
38	45	10.6	77	17	O28902	O28902 archaeglob
39	45	10.6	78	16	Q9K1Q4	Q9klq4 neisseria m
40	45	10.6	79	10	Q9LUX7	Q9lux7 pyrus pyrif
41	45	10.6	80	6	Q95MH0	Q95mh0 macaca mula
42	45	10.6	80	6	Q95MH1	Q95mh1 papio anubi
43	45	10.6	80	6	Q95MH3	Q95mh3 gorilla gor
44	45	10.6	80	6	Q95MH2	Q95mh2 pongo pygma
45	45	10.6	80	6	Q95MH4	Q95mh4 pan troglod
46	45	10.6	80	6	Q95MG9	Q95mg9 macaca sile
47	45	10.6	82	17	O27686	O27686 methanobact
48	44.5	10.5	55	16	Q7UI37	Q7ui37 rhodopirell
49	44.5	10.5	64	16	Q8YQU5	Q8yqu5 anabaena sp
50	44.5	10.5	70	15	Q9J124	Q9j124 human immun
51	44.5	10.5	76	2	Q8VN35	Q8vn35 helicobacte
52	44.5	10.5	76	2	Q8VN29	Q8vn29 helicobacte
53	44.5	10.5	77	16	Q7UL23	Q7ul23 rhodopirell
54	44.5	10.5	79	16	P73181	P73181 synechocyst
55	44.5	10.5	81	12	Q7TE81	Q7te81 dolichos ye
56	44.5	10.5	81	17	Q8TVM9	Q8tvm9 methanopyru
57	44	10.4	60	17	Q8TTK0	Q8ttk0 methanosarc

58	44	10.4	63	5	O16833	O16833 drosophila
59	44	10.4	65	15	Q9J136	Q9j136 human immun
60	44	10.4	68	15	Q9J162	Q9j162 human immun
61	44	10.4	72	15	Q9J156	Q9j156 human immun
62	44	10.4	78	16	Q9RUR5	Q9rur5 deinococcus
63	44	10.4	78	16	Q8X5B5	Q8x5b5 escherichia
64	44	10.4	85	5	Q9VHZ4	Q9vhz4 drosophila
65	43.5	10.3	50	16	Q89TN7	Q89tn7 bradyrhizob
66	43.5	10.3	61	15	Q97614	Q97614 human immun
67	43.5	10.3	67	16	Q82DZ6	Q82dz6 streptomyce
68	43.5	10.3	67	16	Q7USP0	Q7usp0 rhodopirell
69	43.5	10.3	68	12	Q9YRD4	Q9yrd4 largemouth
70	43.5	10.3	69	15	Q9WMQ6	Q9wmq6 human immun
71	43.5	10.3	74	16	Q7UXK2	Q7uxk2 rhodopirell
72	43.5	10.3	76	2	Q8VN31	Q8vn31 helicobacte
73	43.5	10.3	80	3	Q9HGR8	Q9hgr8 choanephora
74	43.5	10.3	83	16	Q8XAC1	Q8xac1 escherichia
75	43.5	10.3	84	9	Q8SC65	Q8sc65 stx2 conver
76	43	10.2	58	9	Q9MC73	Q9mc73 bacterioph
77	43	10.2	65	15	Q9J158	Q9j158 human immun
78	43	10.2	68	11	Q8BU24	Q8bu24 mus musculu
79	43	10.2	68	15	Q9J183	Q9j183 human immun
80	43	10.2	69	15	O90585	O90585 human immun
81	43	10.2	69	15	Q9J169	Q9j169 human immun
82	43	10.2	70	15	Q97583	Q97583 human immun
83	43	10.2	72	15	Q97596	Q97596 human immun
84	43	10.2	73	6	Q8MJD6	Q8mjd6 sus scrofa
85	43	10.2	73	16	Q8VIY7	Q8viy7 mycobacteri
86	43	10.2	76	16	Q835F1	Q835f1 enterococcu
87	43	10.2	83	2	Q7WX29	Q7wx29 alcaligenes
88	43	10.2	83	16	Q82UD0	Q82ud0 nitrosomona
89	42.5	10.0	49	16	Q82NT1	Q82nt1 streptomyce
90	42.5	10.0	52	16	Q9PDF0	Q9pdf0 xylella fas
91	42.5	10.0	56	2	Q9KK61	Q9kk61 mycobacteri
92	42.5	10.0	64	16	Q8ZS74	Q8zs74 anabaena sp
93	42.5	10.0	66	16	Q8NWI6	Q8nwi6 staphylococ
94	42.5	10.0	68	9	O80094	O80094 staphylococ
95	42.5	10.0	68	15	Q97588	Q97588 human immun
96	42.5	10.0	74	2	Q49718	Q49718 mycobacteri
97	42.5	10.0	74	16	Q8PlQ2	Q8plq2 streptococc
98	42.5	10.0	76	10	Q8LCZ5	Q8lcz5 arabidopsis
99	42.5	10.0	78	16	Q89S16	Q89s16 bradyrhizob
100	42.5	10.0	80	16	Q8E852	Q8e852 shewanella

ALIGNMENTS

RESULT 1

Q93J23

ID Q93J23 PRELIMINARY; PRT; 67 AA.

AC Q93J23;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein SCO3984.

GN SCO3984 OR SCBAC25E3.21.

RESULT 2

Q7VBG2

ID Q7VBG2 PRELIMINARY; PRT; 78 AA.
AC Q7VBG2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Predicted protein.
GN PRO1133.
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=1219;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SARG / CCMP 1375 / SS120;
RX MEDLINE=22810154; PubMed=12917486;
RA Dufresne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,
RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,
RA Makarova K.S., Ostrowski M., Oztas S., Robert C., Rogozin I.B.,
RA Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,
RA Wolf Y.I., Hess W.R.;
RT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,
RT a nearly minimal oxyphototrophic genome."
RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).
DR EMBL; AE017164; AAQ00178.1; -.
KW Complete proteome.
SQ SEQUENCE 78 AA; 8555 MW; 338B0D6AE8B40155 CRC64;

Query Match 13.5%; Score 57; DB 16; Length 78;
Best Local Similarity 35.1%; Pred. No. 52;
Matches 13; Conservative 5; Mismatches 17; Indels 2; Gaps 1;

Qy 22 LVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCL 58
|| || | |:: | |:: | | ||
Db 19 LVGMD--GHPHPVLDTPYESVDAAIGAAGQWTSKHCL 53

RESULT 3

O80316

ID O80316 PRELIMINARY; PRT; 58 AA.
AC O80316;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Orf52 (Fragment).
GN H.
OS Bacteriophage 186.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC P2-like viruses.
OX NCBI_TaxID=29252;
RN [1]
RP SEQUENCE FROM N.A.
RA Xue Q.;
RT "Studies on the tail region of the temperate coliphage 186 genome."
RL Thesis (1993), University of Adelaide.

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98371265; PubMed=9705261;
 RA Portelli R., Dodd I.B., Xue Q., Egan J.B.;
 RT "The late-expressed region of the temperate coliphage 186 genome."
 RL Virology 248:117-130(1998).
 DR EMBL; U32222; AAC34169.1; -.
 FT NON_TER 1 1
 FT VARIANT 15 15 S -> *.
 FT VARIANT 51 51 Q -> *.
 SQ SEQUENCE 58 AA; 6491 MW; 1199113D8CDEB8E6 CRC64;

 Query Match 13.2%; Score 56; DB 9; Length 58;
 Best Local Similarity 38.5%; Pred. No. 47;
 Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy 38 PTEALSVAVEEGLAWRKKGCLRLGTH 63
 | : | | : : | : | : | | |
 Db 31 PSELYSLSLTELITWREKALQSGNH 56

RESULT 4

Q8IRX0

ID Q8IRX0 PRELIMINARY; PRT; 75 AA.
 AC Q8IRX0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE CG32806-PB.
 GN CG32806.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanenvong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AE003422; AAN09059.1; -.
 DR FlyBase; FBgn0052806; CG32806.

SQ SEQUENCE 75 AA; 8466 MW; ED0FDFC83591E05C CRC64;

Query Match 13.0%; Score 55; DB 5; Length 75;
Best Local Similarity 29.3%; Pred. No. 85;
Matches 17; Conservative 10; Mismatches 27; Indels 4; Gaps 3;

Qy 5 GCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEE-GLAWRKKGCLRLG 61
|| : || ||: : :: | : | ||: || || : | | : |
Db 6 GCRGLAKSPRRSVCD-EMISRDALPARVAPSEMPKPEVATEEPSVQW--NACYWIG 60

RESULT 5

Q89Y84

ID Q89Y84 PRELIMINARY; PRT; 73 AA.
AC Q89Y84;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Bsr0071 protein.
GN BSR0071.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005935; BAC45336.1; -.
KW Complete proteome.
SQ SEQUENCE 73 AA; 8063 MW; C4CA103399C8734B CRC64;

Query Match 12.8%; Score 54; DB 16; Length 73;
Best Local Similarity 34.9%; Pred. No. 1.1e+02;
Matches 15; Conservative 8; Mismatches 12; Indels 8; Gaps 2;

Qy 47 EEGLAWRKKGCLRLG----THGSPTASSQSSAT----NMAIHR 81
| | | |::| ||:| | | ::::| | : |||
Db 11 ESGWATRREGALRVGSTHHTQAEATEAARSTALREHGEVVIHR 53

RESULT 6

Q23341

ID Q23341 PRELIMINARY; PRT; 80 AA.
AC Q23341;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN ZC477.4.

OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 RT investigating biology. The *C. elegans* Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z.;
 RT "The sequence of *C. elegans* cosmid ZC477.";
 RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; U40802; AAK19010.1; -.
 DR PIR; T27603; T27603.
 DR WormPep; ZC477.4; CE05060.
 KW Hypothetical protein.
 SQ SEQUENCE 80 AA; 8481 MW; AE43A8268EB6C423 CRC64;

Query Match 12.8%; Score 54; DB 5; Length 80;
 Best Local Similarity 28.4%; Pred. No. 1.2e+02;
 Matches 23; Conservative 9; Mismatches 35; Indels 14; Gaps 2;

Qy 6 CSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGS 65
 || || || || :| | | | | :| | | :| | | :| | |
 Db 4 CSPLKILPGASSSSSSSTA-----SSQIRPPSLSLASLSEELRVEECGSPRVGAKES 56
 Qy 66 -----PTASSQSSATNMAI 79
 | || | : :
 Db 57 SFYCTEQPAQSSYSREDKLCL 77

RESULT 7

Q9JMT2

ID Q9JMT2 PRELIMINARY; PRT; 84 AA.
 AC Q9JMT2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE YbgA protein.
 GN YBGA.
 OS *Escherichia coli*.
 OG Plasmid F.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;
 RT "Complete nucleotide sequence of the F plasmid: Its implications for
 RT organization and diversification of plasmid genomes.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90317835; PubMed=2164585;
 RA Yoshioka Y., Fujita Y., Ohtsubo E.;
 RT "Nucleotide sequence of the promoter-distal region of the tra operon
 RT of plasmid R100, including traI (DNA helicase I) and traD genes.";
 RL J. Mol. Biol. 214:39-53(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87194554; PubMed=3032897;
 RA Saadi S., Maas W.K., Hill D.F., Bergquist P.L.;
 RT "Nucleotide sequence analysis of RepFIC, a basic replicon present in
 RT IncFI plasmids P307 and F, and its relation to the RepA replicon of
 RT IncFII plasmids.";
 RL J. Bacteriol. 169:1836-1846(1987).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95337425; PubMed=7612932;
 RA Broom J.E., Hill D.F., Hughes G., Jones W.A., McNaughton J.C.,
 RA Stockwell P.A., Petersen G.B.;
 RT "Sequence of a transposon identified as Tn1000 (gamma delta).";
 RL DNA Seq. 5:185-189(1995).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Eichenlaub R.;
 RT "F Plasmid DNA complete mini-F region (F coordinates 40.301F to
 RT 49.869F).";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86139869; PubMed=3949712;
 RA Helsberg M., Eichenlaub R.;
 RT "Twelve 43-base-pair repeats map in a cis-acting region essential for
 RT partition of plasmid mini-F.";
 RL J. Bacteriol. 165:1043-1045(1986).
 RN [7]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99296678; PubMed=10366527;
 RA Manwaring N.P., Skurray R.A., Firth N.;
 RT "Nucleotide sequence of the F plasmid leading region.";
 RL Plasmid 41:219-225(1999).
 RN [8]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94359430; PubMed=7915817;
 RA Frost L.S., Ippen-Ihler K., Skurray R.A.;
 RT "An analysis of the sequence and gene products of the transfer region
 RT of the F sex factor.";
 RL Microbiol. Rev. 58:162-210(1994).
 DR EMBL; AP001918; BAA97888.1; -.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 KW Plasmid; Complete proteome.

SQ SEQUENCE 84 AA; 9265 MW; 183C60CAF87121F7 CRC64;

Query Match 12.8%; Score 54; DB 16; Length 84;
Best Local Similarity 38.2%; Pred. No. 1.3e+02;
Matches 13; Conservative 2; Mismatches 19; Indels 0; Gaps 0;

Qy 51 AWRKKGCLRLGTHGSPTASSQSSATNMAIHRSQP 84
| | ||: | || | | : | | |
Db 43 AMRAGGCIHPSGRWCPVASSTVPATGLHQHSDP 76

RESULT 8

Q91LH8

ID Q91LH8 PRELIMINARY; PRT; 80 AA.
AC Q91LH8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF62 (Wsv087) (WSSV144).
OS White spot syndrome virus (WSSV).
OC Viruses; dsDNA viruses, no RNA stage; Nimaviridae.
OX NCBI_TaxID=92652;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21342572; PubMed=11448154;
RA van Hulten M.C.W., Witteveldt J., Peters S., Kloosterboer N.,
RA Tarchini R., Fiers M., Sandbrink H., Lankhorst R.K., Vlak J.M.;
RT "The white spot syndrome virus DNA genome sequence.";
RL Virology 286:7-22(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA van Hulten M.C.W., Witteveldt J., Peters S., Kloosterboer N.,
RA Tarchini R., Fiers M., Sandbrink H., Lankhorst R.K., Vlak J.M.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21548311; PubMed=11689662;
RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;
RT "Complete genome sequence of the shrimp white spot bacilliform
RT virus.";
RL J. Virol. 75:11811-11820(2001).
RN [4]
RP SEQUENCE FROM N.A.
RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Taiwan;
RX MEDLINE=20517548; PubMed=11062040;
RA Tsai M.F., Yu H.T., Tzeng H.F., Leu J.H., Chou C.M., Huang C.J.,
RA Wang C.H., Lin J.Y., Kou G.H., Lo C.F.;
RT "Identification and characterization of a shrimp white spot syndrome
RT virus (WSSV) gene that encodes a novel chimeric polypeptide of
RT cellular-type thymidine kinase and thymidylate kinase.";
RL Virology 277:100-110(2000).
RN [6]
RP SEQUENCE FROM N.A.

```

RC      STRAIN=Taiwan;
RX      MEDLINE=21844071; PubMed=11853398;
RA      Chen L.L., Leu J.H., Huang C.J., Chou C.M., Chen S.M., Wang C.H.,
RA      Lo C.F., Kou G.H.;
RT      "Identification of a nucleocapsid protein (VP35) gene of shrimp white
RT      spot syndrome virus and characterization of the motif important for
RT      targeting VP35 to the nuclei of transfected insect cells.";
RL      Virology 293:44-53(2002).
RN      [7]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Taiwan;
RA      Lo C.-F., Kou G.-H.;
RL      Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AF369029; AAK77731.1; -.
DR      EMBL; AF332093; AAL33091.1; -.
DR      EMBL; AF440570; AAL89012.1; -.
SQ      SEQUENCE      80 AA;  8806 MW;  92462B3C00342FB1 CRC64;

```

RESULT 9

```
ID      Q854J6          PRELIMINARY;      PRT;      82 AA.
AC      Q854J6;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      Gp68.
OS      Mycobacteriophage Omega.
OC      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX      NCBI_TaxID=205879;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22592660; PubMed=12705866;
RA      Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,
RA      Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R.,
RA      Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,
RA      Kriakov J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,
RA      Hatfull G.F.;
RT      "Origins of highly mosaic mycobacteriophage genomes.";
RL      Cell 113:171-182(2003).
DR      EMBL; AY129338; AAN12712.1; -.
SQ      SEQUENCE      82 AA;  9176 MW;  8E9859C285BE0942 CRC64;
```

Matches 14; Conservative 7; Mismatches 23; Indels 4; Gaps 1;

```
Qy      40 EALSVAVEEGLAWRKKGCLRLGT----HGSPTASSQSSATNMAIHR SQ 83
          | | | |: || | :|: | :| : | :|| :
Db      33 EMLGVDVDTVKRWRKNGLELVGSRSLCRGAPVEPMLNVAAASRLHRKE 80
```

RESULT 10

Q853P2

ID Q853P2 PRELIMINARY; PRT; 72 AA.
AC Q853P2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Gp35.
OS Mycobacteriophage Bxz1.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
OX NCBI_TaxID=205877;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22592660; PubMed=12705866;
RA Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,
RA Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R.,
RA Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,
RA Kriakov J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,
RA Hatfull G.F.;
RT "Origins of highly mosaic mycobacteriophage genomes.";
RL Cell 113:171-182(2003).
DR EMBL; AY129337; AAN16695.1; -.
SQ SEQUENCE 72 AA; 8125 MW; 6F1E7F9C0D5400D1 CRC64;

Query Match 11.8%; Score 50; DB 9; Length 72;
Best Local Similarity 37.1%; Pred. No. 3.1e+02;
Matches 13; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

```
Qy      31 KSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGS 65
          | | | || |: | : | :||| | |
Db      18 KGTVSSGKTSALTVRIPESVRVEMKRRVRLGLHKS 52
```

RESULT 11

Q7VCF8

ID Q7VCF8 PRELIMINARY; PRT; 73 AA.
AC Q7VCF8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Predicted protein.
GN PRO0782.
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=1219;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SARG / CCMP 1375 / SS120;
RX MEDLINE=22810154; PubMed=12917486;

ID Q8LRC4 PRELIMINARY; PRT; 64 AA.

AC Q8LRC4;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE P0031D02.18 protein.
 GN P0031D02.18.
 OS *Oryza sativa* (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; *Oryza*.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0031D02.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AP003230; BAB93190.1; -.
 DR Gramene; Q8LRC4; -.
 SQ SEQUENCE 64 AA; 7079 MW; 9651126171640B21 CRC64;

Query Match 11.3%; Score 48; DB 10; Length 64;
 Best Local Similarity 30.3%; Pred. No. 4.6e+02;
 Matches 20; Conservative 7; Mismatches 13; Indels 26; Gaps 4;

Qy 21 SILVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRL---GTHGSPTASSQSSATNM 77
 | : :: | | : | || : | : | || | ||
 Db 16 SIAHLEFP LQ-----ESPISVLSVLGE-----RKSQ LRLQLAGLHGS----- 53

 Qy 78 AIHRSQ 83
 ||| |
 Db 54 -IHREQ 58

RESULT 14

Q9J125

ID Q9J125 PRELIMINARY; PRT; 66 AA.
 AC Q9J125;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Gag protein (Fragment).
 GN GAG.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FIN9399;
 RA Liitsola K., Holmstrom P., Laukkanen T., Brummer-Korvenkontio H.,
 RA Leinikki P., Salminen M.O.;
 RT "Analysis of HIV-1 genetic subtypes in Finland reveals good
 RT correlation between molecular and epidemiological data.";
 RL Scand. J. Infect. Dis. 0:0-0(2000).
 DR EMBL; AF219348; AAF30254.1; -.
 DR HSSP; P05888; 1AAF.

DR GO; GO:0019012; C:virion; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro; IPR001878; Znf_CCHC.
 DR PRINTS; PR00939; C2HCZNFINGER.
 KW Core protein; Polyprotein.
 FT NON_TER 1 1
 FT NON_TER 66 66
 SQ SEQUENCE 66 AA; 7236 MW; F74E42EF9F24AD6E CRC64;

Query Match 11.2%; Score 47.5; DB 15; Length 66;
 Best Local Similarity 35.9%; Pred. No. 5.5e+02;
 Matches 14; Conservative 4; Mismatches 20; Indels 1; Gaps 1;

Qy 26 DFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHG 64
 :| ||: | : | | | |||| : | |
 Db 28 NFKGQR-RXLSASTVAEGHLARNCRAPRKKGCWKCGKEG 65

RESULT 15

Q8MMG0

ID Q8MMG0 PRELIMINARY; PRT; 64 AA.
 AC Q8MMG0;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG30154-PA.
 GN CG30154.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanenvong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AE003791; AAM68386.1; -.
 DR FlyBase; FBgn0050154; CG30154.
 SQ SEQUENCE 64 AA; 7293 MW; DE4431D8199CAB17 CRC64;

Query Match 11.1%; Score 47; DB 5; Length 64;
Best Local Similarity 32.5%; Pred. No. 6.1e+02;
Matches 13; Conservative 4; Mismatches 19; Indels 4; Gaps 2;

Qy 42 LSVAVEEGLAWRKKGCLRLGTHGSPT-ASSQSSATNMAIH 80
||| |:| :| ||| :| |:||
Db 19 LSV---HGMPWKWGPASNSGPHGGPAWNGGEESTDNIVIH 55

RESULT 16

Q8E7B7

ID Q8E7B7 PRELIMINARY; PRT; 68 AA.
AC Q8E7B7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN GBS0238.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766844; CAD45883.1; -.
DR SagaList; gbs0238; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 68 AA; 7450 MW; 33108A42C112BF80 CRC64;

Query Match 11.1%; Score 47; DB 16; Length 68;
Best Local Similarity 32.7%; Pred. No. 6.5e+02;
Matches 16; Conservative 10; Mismatches 15; Indels 8; Gaps 4;

Qy 16 SISENSLVAMDFS-GQKSRVIENPTEALSVAVEEGL-AWRKKGCLRLGT 62
:|::| |:|: || | |:| | | : | :: | |:||
Db 4 TINKNDLIALGFSEGTSKRIIRQGKELL---IARGFRVYQNK---RIGT 46

RESULT 17

Q84Z53

ID Q84Z53 PRELIMINARY; PRT; 72 AA.
AC Q84Z53;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE P0686C03.33 protein.
GN P0686C03.33.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
 RT clone:P0686C03."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AP004761; BAC56794.1; -.
 SQ SEQUENCE 72 AA; 7591 MW; 66495F03E06E6DC1 CRC64;

Query Match 11.1%; Score 47; DB 10; Length 72;
 Best Local Similarity 29.3%; Pred. No. 7e+02;
 Matches 22; Conservative 8; Mismatches 39; Indels 6; Gaps 2;

Qy 6 CSSQSI SPMSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTHGS 65
 | | | : | | : | || : | | : | |
 Db 2 CRCGSGKEQRQGVDSAVCGGQSTMSTAVGFIPT---LALERGSATTRSASLNLAT--K 55
 Qy 66 PTASSQSSATNMAIH 80
 | || : | ||
 Db 56 TTESSPTAHTRYCIH 70

RESULT 18

Q8R0X3

ID Q8R0X3 PRELIMINARY; PRT; 79 AA.
 AC Q8R0X3;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Similar to LOC164714.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thymus;
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC026208; AAH26208.1; -.
 SQ SEQUENCE 79 AA; 8632 MW; 806C30C3455C10BE CRC64;

Query Match 11.1%; Score 47; DB 11; Length 79;
 Best Local Similarity 28.0%; Pred. No. 7.8e+02;
 Matches 14; Conservative 5; Mismatches 11; Indels 20; Gaps 2;

Qy 37 NPTEALSVAVEEG-----LAWRKKGC-----LRLGTHGSP 66
 : | | | : | | | : | | | | | | | |
 Db 16 HPQERLCPSATQGIHAGSLNWRPCTCGTLQTIEFSRQYLHGERLGTGTRGAP 65

RESULT 19

Q8EF85

RC STRAIN=TAP44-3; TRANSPOSON=Tn5044;
 RX MEDLINE=99406912; PubMed=10476039;
 RA Minakhina S., Kholodii G., Mindlin S., Yurieva O., Nikiforov V.;
 RT "Tn5053 family transposons are res site hunters sensing plasmidal res
 RT sites occupied by cognate resolvases.";
 RL Mol. Microbiol. 33:1059-1068(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TAP44-3; TRANSPOSON=Tn5044;
 RA Kholodii G., Yurieva O., Mindlin S., Gorlenko Z., Rybochkin V.,
 RA Nikiforov V.;
 RT "Tn5044, a novel Tn3 family transposon coding for temperature
 RT sensitive mercury resistance.";
 RL Res. Microbiol. 151:1-12(2000).
 DR EMBL; Y17691; CAB65713.1; -.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 79 AA; 8626 MW; 1639B3E026E36706 CRC64;

Query Match 11.0%; Score 46.5; DB 2; Length 79;
 Best Local Similarity 31.0%; Pred. No. 9e+02;
 Matches 18; Conservative 14; Mismatches 15; Indels 11; Gaps 4;

Qy 2 GRSGCSSQSPMRISISNSLVA--MDFSGQKSRVIE-NPTEA-LSVAVEEGLAWRKK 55
 || | : | | : : | : : : : : : || | : || : || |
 Db 28 GRKGDLSRFI-----EEAVRAHILELSAEQAKAVNAHLSEAELTDAVDEALAWASK 78

RESULT 21

Q96U90

ID Q96U90 PRELIMINARY; PRT; 80 AA.
 AC Q96U90;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Probable ribosomal protein S19, mitochondrial.
 GN B1109.070.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AL513409; CAD11378.1; -.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR002222; Ribosomal_S19.
 DR Pfam; PF00203; Ribosomal_S19; 1.

Query Match 11.0%; Score 46.5; DB 3; Length 80;
Best Local Similarity 25.0%; Pred. No. 9.1e+02;
Matches 12; Conservative 14; Mismatches 21; Indels 1; Gaps 1;

RESULT 22

```

ID      Q7V3F4          PRELIMINARY;          PRT;          80 AA.
AC      Q7V3F4;
DT      01-OCT-2003 (TrEMBLrel. 25, Created)
DT      01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Hypothetical protein.
GN      PMM0121.
OS      Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
OC      Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC      Prochlorococcus.
OX      NCBI_TaxID=59919;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22825698; PubMed=12917642;
RA      Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA      Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA      Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
RA      Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA      Webb E.A., Zinser E.R., Chisholm S.W.;
RT      "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT      niche differentiation.";
RL      Nature 424:1042-1047(2003).
DR      EMBL; BX572090; CAE18580.1; -.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE      80 AA;  9218 MW;  19A642863632D7CA CRC64;

```

```
QY      8 SQSISPMRSISENSLVAMD---FSGQ-----KSRVIENPTEALSVAVEEG-LAWR 53
      ::||| : : : ||| :| : | : | : | : | : | : | : | : | : |
Db      2 TESI-PKKPLKKGSLVFDIKSIYDGSVEALASDQDLPYSYIFEGPGEILSIKEEYAQVRWR 60
```

Qy	54	K	54
		:	
Db	61	R	61

RESULT 23

08B6L1

ID Q8B6L1 PRELIMINARY; PRT; 83 AA.
AC Q8B6L1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Coat protein (Fragment).
OS Soybean dwarf virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Luteovirus.
OX NCBI_TaxID=12049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YP;
RA Terauchi H., Honda K., Yamagishi N., Kanematsu S., Ishiguro K.,
RA Hidaka S.;
RT "The N-terminal region of readthrough domain is closely related to
RT aphid transmission specificity of Soybean dwarf virus.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB076045; BAC54080.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001517; Luteo_coat.
DR Pfam; PF00894; Luteo_coat; 1.
DR PRINTS; PR00915; LUTEOGP1COAT.
DR ProDom; PD001068; Luteo_coat; 1.
FT NON_TER 1 1
SQ SEQUENCE 83 AA; 9256 MW; 138B9DD62E136293 CRC64;

Query Match 11.0%; Score 46.5; DB 12; Length 83;
Best Local Similarity 32.0%; Pred. No. 9.5e+02;
Matches 16; Conservative 5; Mismatches 24; Indels 5; Gaps 1;

Qy 4 SGCSSQSI SPMR SISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 53
|| : : | || : || || || : : || ||
Db 4 SGSIAYELDPHCKYSEIQSLNKF SITKSGSKRFPTRAIN-----GLEWR 48

RESULT 24

Q8B6K9

ID Q8B6K9 PRELIMINARY; PRT; 83 AA.
AC Q8B6K9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Coat protein (Fragment).
OS Soybean dwarf virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Luteovirus.
OX NCBI_TaxID=12049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YP;
RA Terauchi H., Honda K., Yamagishi N., Kanematsu S., Ishiguro K.,
RA Hidaka S.;
RT "The N-terminal region of readthrough domain is closely related to
RT aphid transmission specificity of Soybean dwarf virus.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AB076046; BAC54082.1; -.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR001517; Luteo_coat.
 DR Pfam; PF00894; Luteo_coat; 1.
 DR PRINTS; PR00915; LUTEOGP1COAT.
 DR ProDom; PD001068; Luteo_coat; 1.
 FT NON_TER 1 1
 SQ SEQUENCE 83 AA; 9256 MW; 138B9DD62E136293 CRC64;

Query Match 11.0%; Score 46.5; DB 12; Length 83;
 Best Local Similarity 32.0%; Pred. No. 9.5e+02;
 Matches 16; Conservative 5; Mismatches 24; Indels 5; Gaps 1;

Qy 4 SGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 53
 || : : | || : || || || |:: || ||
 Db 4 SGSIAYELDPHCKYSEIQSLNKFISITKSGSKRFPTRAIN-----GLEWR 48

RESULT 25

Q8EIV8

ID Q8EIV8 PRELIMINARY; PRT; 83 AA.
 AC Q8EIV8;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Conserved hypothetical protein.
 GN SO0721.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AE015517; AAN53799.1; -.
 DR TIGR; SO0721; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 83 AA; 9075 MW; AC5D08F38ACB345C CRC64;

Query Match 11.0%; Score 46.5; DB 16; Length 83;
 Best Local Similarity 25.5%; Pred. No. 9.5e+02;
 Matches 13; Conservative 11; Mismatches 16; Indels 11; Gaps 1;

Qy 15 RSISENSLVAMDFSGQ-----KSRVIENPTEALSVAVEEGLAWRK 54

Db : : : : | | : | | | | : | : : | | : | | | : |
23 QALTDNPLMAMGIIGQLGIPPEKLQQLMALVMQNPALIKEAVLELGLDFAK 73

Search completed: July 8, 2004, 08:22:54
Job time : 76.7559 secs

OM protein - protein search, using sw model

Run on: July 8, 2004, 08:03:43 ; Search time 26.4567 Seconds
(without alignments)
165.323 Million cell updates/sec

Title: US-09-936-697-6
Perfect score: 423
Sequence: 1 QGRSGCSSQSISPMRSISEN.....SPTASSQCSSATNMAIHRSQP 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 11046

Minimum DB seq length: 0
Maximum DB seq length: 85

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	51.5	12.2	73	1	RPON	METJA	Q57649 methanococc
2	47	11.1	71	1	Y16K	BPT4	P39243 bacterioph
3	46.5	11.0	62	1	YZ05	METJA	Q60262 methanococc
4	45	10.6	76	1	CD24	RAT	Q07490 rattus norv
5	45	10.6	79	1	MT2	MALDO	O24058 malus domes
6	45	10.6	82	1	RADC	STAAU	P31337 staphylococ
7	44.5	10.5	68	1	GNGI	HUMAN	Q9y3k8 homo sapien
8	43	10.2	81	1	PSK6	ARATH	Q81a14 arabidopsis
9	43	10.2	82	1	Y567	METJA	Q57987 methanococc
10	42.5	10.0	67	1	HFO2	METFO	P48783 methanobact
11	42	9.9	45	1	ATI2	HSVE4	Q00041 equine herp
12	42	9.9	60	1	YA87	STRMU	Q8du62 streptococc
13	42	9.9	78	1	YHGG	ECOLI	P46845 escherichia
14	42	9.9	83	1	TMOB	PSEME	Q00457 pseudomonas
15	41.5	9.8	67	1	HFOB	METFO	P48784 methanobact
16	41	9.7	66	1	RPON	SULSO	Q980z8 sulfolobus
17	41	9.7	79	1	DC13	HUMAN	Q9nnp2 homo sapien

18	40.5	9.6	43	1	CC3_CARCN	P32956	carica cand
19	40.5	9.6	68	1	GBG5_HUMAN	P30670	homo sapien
20	40	9.5	41	1	BAXC_HUMAN	Q07815	homo sapien
21	40	9.5	56	1	HS2M_LYCES	P81161	lycopersico
22	40	9.5	72	1	RPON_THEAC	Q9hl09	thermoplasm
23	40	9.5	72	1	RPON_THEVO	Q979k0	thermoplasm
24	39.5	9.3	35	1	PBP_HYACE	P34175	hyalophora
25	39.5	9.3	60	1	Y574_LACLA	Q9chz4	lactococcus
26	39.5	9.3	83	1	V187_BPT7	P03788	bacterioph
27	39	9.2	43	1	CC4_CARCN	P32957	carica cand
28	39	9.2	45	1	RS22_ECOLI	P28690	escherichia
29	39	9.2	68	1	BRH2_HUMAN	Q9ny43	homo sapien
30	39	9.2	85	1	R37A_MYXGL	Q9y0h7	myxine glut
31	38.5	9.1	62	1	4OT_COMTE	Q9rhm8	comamonas t
32	38.5	9.1	67	1	CSPF_STRCO	P48859	streptomyce
33	38.5	9.1	72	1	YO03_ARCFU	O30268	archaeoglob
34	38.5	9.1	74	1	CO5A_BOVIN	P12082	bos taurus
35	38.5	9.1	74	1	NIFH_NOSSN	P52336	nostoc sp.
36	38.5	9.1	78	1	CINA_STRGV	P29827	streptover
37	38.5	9.1	82	1	S6B1_YEAST	P52870	saccharomyc
38	38	9.0	54	1	IOVO_DRONO	P05560	dromaius no
39	38	9.0	58	1	NINF_BPP22	Q38666	bacterioph
40	38	9.0	58	1	SINI_BACLI	P22755	bacillus li
41	38	9.0	59	1	FER_METBA	P00202	methanosarc
42	38	9.0	63	1	FER2_DESVM	P10624	desulfovibr
43	38	9.0	67	1	YDFZ_ECOLI	P76153	escherichia
44	37	8.7	53	1	LECA_LATAP	P07441	lathyrus ap
45	37	8.7	54	1	IOVO_CASCA	P05559	casuarinus c
46	37	8.7	60	1	NXS1_DENVI	P01418	dendroaspis
47	37	8.7	62	1	4OT3_PSEPU	Q9z431	pseudomonas
48	37	8.7	66	1	CSP7_STRCL	Q01761	streptomyce
49	37	8.7	67	1	HMT1_METTH	P50483	methanobact
50	37	8.7	71	1	EX7S_STRAS	Q8e6m0	streptococc
51	37	8.7	76	1	IPKG_MOUSE	O70139	mus musculu
52	36.5	8.6	80	1	Y509_ECO57	P58092	escherichia
53	36.5	8.6	84	1	RL23_HALMA	P12732	haloarcula
54	36	8.5	55	1	FER_BUTME	P14073	butyribacte
55	36	8.5	59	1	YH13_ARCFU	O28560	archaeoglob
56	36	8.5	60	1	NXS1_DENJA	P01417	dendroaspis
57	36	8.5	62	1	4OT_PSEFL	Q8krr5	pseudomonas
58	36	8.5	63	1	COXO_HUMAN	P15954	homo sapien
59	36	8.5	63	1	COXO_PANTR	P60025	pan troglod
60	36	8.5	74	1	RS18_CHLTE	Q8kam3	chlorobium
61	36	8.5	74	1	SR14_MACRA	O18881	macaca radi
62	36	8.5	76	1	IPKG_HUMAN	Q9y2b9	homo sapien
63	36	8.5	80	1	GCH1_MUCHA	P51598	mucuna hass
64	36	8.5	82	1	RS16_VIBCH	Q9kug0	vibrio chol
65	36	8.5	83	1	TRBG_ECOLI	P41072	escherichia
66	36	8.5	84	1	SCX2_CENNO	P01495	centruroid
67	36	8.5	85	1	NEU1_PAPHA	P32005	papio hamad
68	35.5	8.4	61	1	Y083_ARCFU	O30153	archaeoglob
69	35.5	8.4	63	1	BD02_RAT	O88514	rattus norv
70	35.5	8.4	69	1	GBGU_BOVIN	P50154	bos taurus
71	35.5	8.4	71	1	MT1_CASGL	Q39511	casuarina g
72	35.5	8.4	75	1	ATP9_PARTE	P16001	paramecium
73	35.5	8.4	77	1	IM08_ARATH	Q9xgy4	arabidopsis
74	35.5	8.4	80	1	PSAC_MASLA	O07112	mastigoclad

75	35	8.3	40	1	VIT_MELGA	P56531	meleagris g
76	35	8.3	50	1	RL40_AERPE	Q9yfy7	aeropyrum p
77	35	8.3	52	1	RL40_LEIMA	Q05551	leishmania
78	35	8.3	63	1	CX5A_CONPU	Q9u6z6	conus purpu
79	35	8.3	67	1	HMT2_METTH	O27731	methanobact
80	35	8.3	70	1	ICIC_HIRME	P01051	hirudo medi
81	35	8.3	72	1	RL29_TREPA	O83227	treponema p
82	35	8.3	73	1	CATZ_BOVIN	P05689	bos taurus
83	35	8.3	74	1	CT17_HUMAN	Q9nre2	homo sapien
84	35	8.3	75	1	ME10_EUPRA	P12350	euplotes ra
85	35	8.3	76	1	BB11_SCHCO	P78742	schizophyll
86	35	8.3	77	1	YCXB_CYAPA	P48332	cyanophora
87	35	8.3	79	1	NSGX_HUMAN	Q9uh64	homo sapien
88	35	8.3	81	1	YH25_XYLFA	Q9pcq3	xylella fas
89	35	8.3	82	1	CUD5_SCHGR	P56562	schistocerc
90	35	8.3	85	1	SCX6_CENLL	Q7zlk5	centruoide
91	34.5	8.2	43	1	MUTI_ENTMU	P80925	enterococcu
92	34.5	8.2	55	1	RPON_METTH	O26147	methanobact
93	34.5	8.2	63	1	BD03_MOUSE	Q9wtl0	mus musculu
94	34.5	8.2	71	1	ACAI_ACALU	P81592	acalolepta
95	34.5	8.2	72	1	YHDL_ECOLI	P36675	escherichia
96	34.5	8.2	73	1	RL24_HELPY	P56049	helicobacte
97	34.5	8.2	80	1	MT1_COFAR	P43396	coffea arab
98	34.5	8.2	80	1	VPU_HVIMA	P05924	human immun
99	34.5	8.2	82	1	YHYD_ANACY	P16420	anabaena cy
100	34.5	8.2	85	1	RL10_SERMA	P41192	serratia ma

ALIGNMENTS

RESULT 1

RPON_METJA

ID RPON_METJA STANDARD; PRT; 73 AA.
AC Q57649;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).
GN RPON OR MJ0196.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
 CC of DNA into RNA using the four ribonucleoside triphosphates as
 CC substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- SIMILARITY: Belongs to the archaeal rpoN / eukaryotic RPB10 RNA
 CC polymerase subunit family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67475; AAB98176.1; -.
 DR HSSP; O26147; 1EF4.
 DR TIGR; MJ0196; -.
 DR HAMAP; MF_00250; -, 1.
 DR InterPro; IPR000268; RNA_pol_N.
 DR Pfam; PF01194; RNA_pol_N; 1.
 DR ProDom; PD006539; RNA_pol_N; 1.
 DR PROSITE; PS01112; RNA_POL_N_8KD; 1.
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Metal-binding; Complete proteome.
 FT METAL 7 7 ZINC (BY SIMILARITY).
 FT METAL 10 10 ZINC (BY SIMILARITY).
 FT METAL 44 44 ZINC (BY SIMILARITY).
 FT METAL 45 45 ZINC (BY SIMILARITY).
 SQ SEQUENCE 73 AA; 8695 MW; E716EA406D65B831 CRC64;

Query Match 12.2%; Score 51.5; DB 1; Length 73;
 Best Local Similarity 30.0%; Pred. No. 36;
 Matches 15; Conservative 10; Mismatches 18; Indels 7; Gaps 2;

Qy 13 PMRSISENSLVAMDFSGQKSRVI--ENPTEALSVAVEEGLAWRKKGCLRL 60
 |:| | ::| | | |:: ||| :| :| :| | |:
 Db 4 PIRCFSCGNVIAEVFEEYKERILKGENPKDVL-----DDLGIKKYCCRRM 48

RESULT 2

Y16K_BPT4

ID Y16K_BPT4 STANDARD; PRT; 71 AA.
 AC P39243;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical 8.1 kDa protein in ndd-denB intergenic region.
 GN Y16K OR NDD.1.
 OS Bacteriophage T4.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC T4-like viruses.
 OX NCBI_TaxID=10665;
 RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=22514363; PubMed=12626685;
RA Miller E.S., Kutter E., Mosig G., Arisaka F., Kunisawa T., Ruger W.;
RT "Bacteriophage T4 genome.";
RL Microbiol. Mol. Biol. Rev. 67:86-156(2003).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF158101; AAD42616.1; -.
KW Hypothetical protein.
SQ SEQUENCE 71 AA; 8143 MW; 5D56546D2FADAF0C CRC64;

Query Match 11.1%; Score 47; DB 1; Length 71;
Best Local Similarity 36.1%; Pred. No. 1.1e+02;
Matches 13; Conservative 5; Mismatches 16; Indels 2; Gaps 1;

QY 11 ISPMRSISENSLVAMDFSGQKSR--VIENPTEALSV 44
|||::| || | : : ||| || :|
Db 24 ISPLKSTSEKMTV NATLANNSNERFCIENDTET YTV 59

RESULT 3

YZ05_METJA

ID YZ05_METJA STANDARD; PRT; 62 AA.
AC Q60262;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJECL05.
GN MJECL05.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; L77118; AAC37071.1; -.
 DR PIR; E64510; E64510.
 DR TIGR; MJECL05; -.
 KW Hypothetical protein; Complete proteome.
 FT DOMAIN 3 15 ILE-RICH.
 SQ SEQUENCE 62 AA; 7327 MW; 1624EC72E75EBAD7 CRC64;

Query Match 11.0%; Score 46.5; DB 1; Length 62;
 Best Local Similarity 28.6%; Pred. No. 1.1e+02;
 Matches 12; Conservative 8; Mismatches 21; Indels 1; Gaps 1;

QY 15 RSISENSLVAMDFS-GQKSRVIENPTEALSVAVEEGLAWRKK 55
 : ::| | :: | | : | | :|||: | ||
 Db 18 KKVAERFLKDLESSQGMWKEIRERAERAKKQLEEGIEWAKK 59

RESULT 4

CD24_RAT

ID CD24_RAT STANDARD; PRT; 76 AA.
 AC Q07490;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Signal transducer CD24 precursor (Heat stable antigen) (HSA)
 DE (Nectadrin).
 GN CD24A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Embryonic brain;
 RX MEDLINE=94122434; PubMed=8292828;
 RA Shirasawa T., Akashi T., Sakamoto K., Takahashi H., Maruyama N.,
 RA Hirokawa K.;
 RT "Gene expression of CD24 core peptide molecule in developing brain
 RT and developing non-neural tissues.";
 RL Dev. Dyn. 198:1-13(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fischer;
 RX MEDLINE=97157759; PubMed=9004038;
 RA Magnaldo T.A., Barrandon Y.;
 RT "CD24 (heat stable antigen, nectadrin), a novel keratinocyte
 RT differentiation marker, is preferentially expressed in areas of the
 RT hair follicle containing the colony-forming cells.";
 RL J. Cell Sci. 109:3035-3045(1996).
 CC -!- FUNCTION: May have a pivotal role in cell differentiation. The
 CC triggering mechanism of signal transduction may be due to the

CC interactions of differentiating cells with the matrix substrate
 CC via the carbohydrate structure of the molecule. In this way, the
 CC signal transducer can play very different roles in different cell
 CC types as a direct consequence of its glycosylation.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- TISSUE SPECIFICITY: Expressed in the central nervous system, in
 CC postmitotic cells of spinal cord, hindbrain, midbrain and
 CC forebrain. Expressed in epithelium during the development of non-
 CC neural tissues. Expressed in tooth development, specifically in
 CC mesenchymal cells differentiating into odontoblast in dental
 CC papilla, as well as in the developing eye and hair follicle.
 CC -!- DEVELOPMENTAL STAGE: Detected in primitive ectoderm, mesoderm and
 CC ventral endoderm; down-regulated when organogenesis is completed.
 CC -!- PTM: Extensively O-glycosylated (By similarity). The carbohydrate
 CC structure may be regulated in a tissue-specific and developmental
 CC stage-specific manner.
 CC -!- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z11663; CAA77731.1; -.
 DR EMBL; U49062; AAA91470.1; -.
 DR PIR; I53107; I53107.
 KW Glycoprotein; GPI-anchor; Membrane; Signal; Differentiation;
 KW Lipoprotein.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 56 SIGNAL TRANSDUCER CD24.
 FT PROPEP 57 76 REMOVED IN MATURE FORM (BY SIMILARITY).
 FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT LIPID 56 56 GPI-anchor amidated serine (Potential).
 SQ SEQUENCE 76 AA; 7862 MW; 42846E70EC39D958 CRC64;

Query Match 10.6%; Score 45; DB 1; Length 76;
 Best Local Similarity 24.4%; Pred. No. 2.1e+02;
 Matches 19; Conservative 10; Mismatches 15; Indels 34; Gaps 4;

Qy 6 CSSQSISPMRSISENSLVAMDFSGQKS-RVIENPTEALSVAVEEGLAWRKKGCLRLGTHG 64
 | : | : : | | | : | | | | : : ||
 Db 26 CNQTSVAP-----FSGNQSISAAPNPTNATT-----RSGC----- 55
 Qy 65 SPTASSQSSATNMAIHRS 82
 : | | : | : | : |
 Db 56 ---SSLQSTAGLLALSLS 70

RESULT 5
 MT2_MALDO
 ID MT2_MALDO STANDARD; PRT; 79 AA.
 AC 024058;

DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Metallothionein-like protein type 2.
 GN MT1.
 OS Malus domestica (Apple) (Malus sylvestris).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Rosales; Rosaceae; Maloideae; Malus.
 OX NCBI_TaxID=3750;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fruit cortical tissue;
 RA Reid S.J., Ross G.S.;
 RT "Up-regulation of two cDNA clones encoding metallothionein-like
 RT proteins in apple fruit during cool storage.";
 RL Physiol. Plantarum 100:183-189(1997).
 CC -!- FUNCTION: Metallothioneins have a high content of cysteine
 CC residues that bind various heavy metals.
 CC -!- SIMILARITY: Belongs to the metallothionein superfamily; family 15.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U61973; AAC23697.1; -.
 DR PIR; T17014; T17014.
 DR InterPro; IPR002400; GF_cysknot.
 DR InterPro; IPR000347; Metallothion_15.
 DR Pfam; PF01439; Metallothio_2; 1.
 DR PRINTS; PR00438; GFCYSKNOT.
 DR ProDom; PD001611; Metallothion_15; 1.
 KW Metal-binding; Metal-thiolate cluster.
 SQ SEQUENCE 79 AA; 7836 MW; 8ADC58B1D8B644CC CRC64;

Query Match 10.6%; Score 45; DB 1; Length 79;
 Best Local Similarity 30.6%; Pred. No. 2.2e+02;
 Matches 15; Conservative 7; Mismatches 21; Indels 6; Gaps 2;

Qy 4 SGCSSQSI SPMRSISENS---LVAMDFSGQKSRVIENPTEALSVAVEEG 49
 |||: ::| | | : | : ||| : : || | |
 Db 19 SGCNGCGMAPDLSYMEGSTTETLVMGVAPQKSHM---EASEMGVAAENG 64

RESULT 6

RADC_STAAU

ID RADC_STAAU STANDARD; PRT; 82 AA.
 AC P31337;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA repair protein radC homolog (25 kDa protein) (Fragment).
 GN RADC.

OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RN450;
 RA Murphy E.;
 RL Submitted (JAN-1986) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A.
 RC STRAIN=RN450;
 RX MEDLINE=84117462; PubMed=6320000;
 RA Murphy E., Loeffdahl S.;
 RT "Transposition of Tn554 does not generate a target duplication.";
 RL Nature 307:292-294(1984).
 CC -!- FUNCTION: Involved in DNA repair (By similarity).
 CC -!- SIMILARITY: Belongs to the radC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; K02985; AAA26680.1; -.
 DR HAMAP; MF_00018; -; 1.
 DR InterPro; IPR001405; RadC.
 DR Pfam; PF04002; RadC; 1.
 DR ProDom; PD007415; RadC; 1.
 DR PROSITE; PS01302; RADC; 1.
 KW DNA repair.
 FT NON_TER 1 1
 SQ SEQUENCE 82 AA; 8920 MW; 65E8BF06E3DEC3A4 CRC64;

Query Match 10.6%; Score 45; DB 1; Length 82;
 Best Local Similarity 27.3%; Pred. No. 2.3e+02;
 Matches 15; Conservative 7; Mismatches 27; Indels 6; Gaps 2;

Qy 27 FSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTH--GSPTASSQSSATNMAI 79
 | | : | : | | : | | : : | | | : | | :
 Db 1 FKGTLNSSIVHPREIFSIIVRE---NANAIIVHNHPSGDVTPSQEDIITTMRL 51

RESULT 7

GNGL_HUMAN

ID GNGL_HUMAN STANDARD; PRT; 68 AA.
 AC Q9Y3K8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-5 like
 DE subunit.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20277477; PubMed=10819326;
 RA Hurowitz E.H., Melnyk J.M., Chen Y.J., Kouros-Mehr H., Simon M.I.,
 RA Shizuya H.;
 RT "Genomic characterization of the human heterotrimeric G protein alpha,
 RT beta, and gamma subunit genes.";
 RL DNA Res. 7:111-120(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Heath P.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are
 CC involved as a modulator or transducer in various transmembrane
 CC signaling systems. The beta and gamma chains are required for the
 CC GTPase activity, for replacement of GDP by GTP, and for G protein-
 CC effector interaction (By similarity).
 CC -!- SUBUNIT: G proteins are composed of 3 units, alpha, beta and
 CC gamma.
 CC -!- SIMILARITY: Belongs to the G protein gamma family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF188178; AAF04568.1; -.
 DR EMBL; AL031319; CAB41647.1; -.
 DR GO; GO:0005576; C:extracellular; NAS.
 DR GO; GO:0003927; F:heterotrimeric G-protein GTPase activity; NAS.
 DR GO; GO:0004871; F:signal transducer activity; NAS.
 DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; NAS.
 DR InterPro; IPR001770; G-gamma.
 DR Pfam; PF00631; G-gamma; 1.
 DR PRINTS; PR00321; GPROTEING.
 DR ProDom; PD003783; G-gamma; 1.
 DR SMART; SM00224; GGL; 1.
 DR PROSITE; PS50058; G_PROTEIN_GAMMA; 1.
 SQ SEQUENCE 68 AA; 7251 MW; 869BCA2A081EAA02 CRC64;

Query Match 10.5%; Score 44.5; DB 1; Length 68;
 Best Local Similarity 38.2%; Pred. No. 2.1e+02;
 Matches 13; Conservative 5; Mismatches 15; Indels 1; Gaps 1;

Qy 43 SVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATN 76
 ||| : | |: ||: | | : ||: ||
 Db 25 SVKVSQAAADLKQFCLQNAQH-DPLLTGVSSSTN 57

RESULT 8
 PSK6_ARATH
 ID PSK6_ARATH STANDARD; PRT; 81 AA.
 AC Q8LA14; Q8W5Q9;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Putative phytosulfokines 6 precursor (AtPSK6) (AtPSK3_2) [Contains:
 DE Phytosulfokine-alpha-like (PSK-alpha-like) (Phytosulfokine-a-like);
 DE Phytosulfokine-beta (PSK-beta) (Phytosulfokine-b)].
 GN PSK6 OR AT3G44735 OR T32N15.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21563059; PubMed=11706167;
 RA Yang H., Matsubayashi Y., Nakamura K., Sakagami Y.;
 RT "Diversity of Arabidopsis genes encoding precursors for
 RT phytosulfokine, a peptide growth factor.";
 RL Plant Physiol. 127:842-851(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
 RA Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
 RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choisine N., Artiguenave F., Robert C., Brottier P.,
 RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
 RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
 RA Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:820-822(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.A.;

RT "Full-length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: Promotes plant cell differentiation, organogenesis and
 CC somatic embryogenesis as well as cell proliferation (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -!- PTM: Sulfation is important for activity and for the binding to a
 CC putative membrane receptor (By similarity).
 CC -!- PTM: PSK-beta is an enzymatic derivative of PSK-alpha
 CC (By similarity).
 CC -!- SIMILARITY: Belongs to the phytosulfokine family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB074573; BAB72177.2; -.
 DR EMBL; AC002534; -; NOT_ANNOTATED_CDS.
 DR EMBL; AY088090; AAM65636.1; -.
 KW Growth factor; Differentiation; Signal; Sulfation; Multigene family.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 72 POTENTIAL.
 FT PEPTIDE 73 77 PHYTOSULFOKINE-ALPHA (POTENTIAL).
 FT PEPTIDE 73 76 PHYTOSULFOKINE-BETA (POTENTIAL).
 FT PROPEP 78 81 POTENTIAL.
 FT MOD_RES 73 73 SULFATION (BY SIMILARITY).
 FT MOD_RES 75 75 SULFATION (BY SIMILARITY).
 FT CONFLICT 4 4 S -> T (IN REF. 3).
 FT CONFLICT 22 22 R -> H (IN REF. 3).
 SQ SEQUENCE 81 AA; 9291 MW; DCCD2A2A08461729 CRC64;

Query Match 10.2%; Score 43; DB 1; Length 81;
 Best Local Similarity 32.4%; Pred. No. 3.8e+02;
 Matches 11; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

Qy 3 RSGCSSQSI\$PMRSI-----SENSLVAMDFSGQ 30
 | | | | : | | | : | : | :
 Db 22 RRGKEDQEINPLVSATSVEEDSVNKLGMGYECGE 55

RESULT 9

Y567_METJA

ID Y567_METJA STANDARD; PRT; 82 AA.
 AC Q57987;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0567.
 GN MJ0567.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschii*.";
 RL Science 273:1058-1073(1996).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U67505; AAB98558.1; -.
 DR PIR; G64370; G64370.
 DR TIGR; MJ0567; -.
 DR InterPro; IPR007167; FeoA.
 DR InterPro; IPR008988; Transcr_rep_C.
 DR Pfam; PF04023; FeoA; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 82 AA; 8766 MW; 3F3810EEFC9F81CE CRC64;

Query Match 10.2%; Score 43; DB 1; Length 82;
 Best Local Similarity 25.9%; Pred. No. 3.8e+02;
 Matches 15; Conservative 10; Mismatches 15; Indels 18; Gaps 3;

Qy 4 SGCSSQSISPMRSISENSLVAMDFS-GQKSRVIEN-----PTEALSVAVEEGLAWR 53
 :|| : ||:| : | :|| | | : :||: ||| :
 Db 20 AGCGAM-----QRLVSMGINIGSKLKVIRNQNGPVIISTKGSNIAIGRGLAMK 67

RESULT 10

HFO2_METFO

ID HFO2_METFO STANDARD; PRT; 67 AA.

AC P48783;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Archaeal histone A2.

GN HFOA2.

OS Methanobacterium formicicum.

OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

OC Methanobacteriaceae; Methanobacterium.

OX NCBI_TaxID=2162;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JF-1;
 RX MEDLINE=95138058; PubMed=7836329;
 RA Darcy T.J., Sandman K.M., Reeve J.N.;
 RT "Methanobacterium formicicum, a mesophilic methanogen, contains three
 RT Hfo histones.";
 RL J. Bacteriol. 177:858-860(1995).
 RN [2]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=95138058; PubMed=7836329;
 RA Sandman K.M., Grayling R.A., Reeve J.N.;
 RL Unpublished results, cited by:
 RL Darcy T.J., Sandman K.M., Reeve J.N.;
 RL J. Bacteriol. 177:858-860(1995).
 CC -!- FUNCTION: Binds and compact DNA (95 to 150 base pairs) to form
 CC nucleosome-like structures that contain positive DNA supercoils.
 CC -!- SUBUNIT: Homodimer or heterodimer (Potential).
 CC -!- SIMILARITY: Belongs to the archaeal histone HMF family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U12931; AAA67722.1; -.
 DR HSSP; P48781; 1B67.
 DR InterPro; IPR003958; CBFA_NFYB_domain.
 DR InterPro; IPR007124; Hist_TAF.
 DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
 KW DNA-binding; Multigene family.
 FT INIT_MET 0 0
 SQ SEQUENCE 67 AA; 7064 MW; 0AAFCAC535BF2E10 CRC64;

Query Match 10.0%; Score 42.5; DB 1; Length 67;
 Best Local Similarity 25.8%; Pred. No. 3.4e+02;
 Matches 16; Conservative 12; Mismatches 25; Indels 9; Gaps 2;

Qy 11 ISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKGCLRLGTH-GSPTAS 69
 |:|: |:|: |: :: |||: |:|| | :| ||| |
 Db 5 IAPVGRIIKNA-----GAQRISDDAKEALAKALEENGEELAKKAVELAKHAGRKTvk 56

Qy 70 SQ 71
 ::
 Db 57 AE 58

RESULT 11
 ATI2_HSVE4
 ID ATI2_HSVE4 STANDARD; PRT; 45 AA.
 AC Q00041;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE Alpha trans-inducing factor 82 kDa protein (Fragment).
 GN 14 OR B7.

OS Equine herpesvirus type 4 (strain 1942) (EHV-4) (Equine herpesvirus
OS type 1 subtype 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10333;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91202570; PubMed=1850013;
RA Whittaker G.R., Riggio M.P., Halliburton I.W., Killington R.A.,
RA Allen G.P., Meredith D.M.;
RT "Antigenic and protein sequence homology between VP13/14, a herpes
RT simplex virus type 1 tegument protein, and gp10, a glycoprotein of
RT equine herpesvirus 1 and 4.";
RL J. Virol. 65:2320-2326(1991).
CC -!- FUNCTION: Modulate alpha trans-inducing factor-dependent
CC activation of alpha genes (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17684; CAA35673.1; -.
DR PIR; S36709; S36709.
DR InterPro; IPR005051; Herpes_UL46.
DR Pfam; PF03387; Herpes_UL46; 1.
KW Transcription regulation; Trans-acting factor.
FT NON_TER 45 45
SQ SEQUENCE 45 AA; 4862 MW; AAE468C9C2B08BE4 CRC64;

Query Match 9.9%; Score 42; DB 1; Length 45;
Best Local Similarity 33.3%; Pred. No. 2.4e+02;
Matches 19; Conservative 4; Mismatches 16; Indels 18; Gaps 3;

Qy 25 MDFSGQKS--RVIENTEALSVAVEEGLAWRKKGCLRLGTHGSPTASSQSSATNMAI 79
|: || | ||: | :| ||| || | |:
Db 1 MEAGSASWARVSKNLIERRAV-----KGCL-----LPTPSDVMDAAVMAL 41

RESULT 12

YA87_STRMU
ID YA87_STRMU STANDARD; PRT; 60 AA.
AC Q8DU62;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable tautomerase SMU.1087 (EC 5.3.2.-).
GN SMU.1087.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=UA159 / ATCC 700610 / Serotype C;
 RX MEDLINE=22295063; PubMed=12397186;
 RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
 RT pathogen."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 CC -!- SIMILARITY: Belongs to the tautomerase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE014946; AAN58785.1; -.
 DR HAMAP; MF_00718; -; 1.
 DR InterPro; IPR004370; Taut.
 DR Pfam; PF01361; Tautomerase; 1.
 DR ProDom; PD404143; Taut; 1.
 KW Isomerase; Complete proteome.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT ACT_SITE 1 1 CATALYTIC BASE (BY SIMILARITY).
 SQ SEQUENCE 60 AA; 6872 MW; 0ADFFDF5985622F4 CRC64;

Query Match 9.9%; Score 42; DB 1; Length 60;
 Best Local Similarity 29.7%; Pred. No. 3.4e+02;
 Matches 19; Conservative 8; Mismatches 17; Indels 20; Gaps 4;

Qy 1 QGRSGCSSQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVE---EGLAW----- 52
 :||| | | | :| | | : | ||: | : || :
 Db 9 EGRS--QEQQIQLAREVTE-----VVSrvAKAPKEAIHVfINDMPEGTYYPHGEM 56
 Qy 53 RKKG 56
 :|||
 Db 57 KKKG 60

RESULT 13

YHGG_ECOLI

ID YHGG_ECOLI STANDARD; PRT; 78 AA.
 AC P46845;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yhgG.
 GN YHGG OR B3410 OR Z4765 OR ECS4252.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U18997; AAA58208.1; -.
 DR EMBL; AE000416; AAC76435.1; -.
 DR EMBL; AE005563; AAG58511.1; -.
 DR EMBL; AP002565; BAB37675.1; -.
 DR PIR; C86006; C86006.
 DR PIR; D91160; D91160.
 DR PIR; E65136; E65136.
 DR EcoGene; EG12933; yhgG.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 78 AA; 8660 MW; 88976DE22CA9024B CRC64;

Query Match 9.9%; Score 42; DB 1; Length 78;
 Best Local Similarity 27.3%; Pred. No. 4.7e+02;
 Matches 15; Conservative 11; Mismatches 21; Indels 8; Gaps 2;

Qy 8 SQSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVE---EGLAWRKKGCLR 59
 ||::: : : | :: |: |: || : : || : |||

RESULT 14

TMOB_PSEME

ID TMOB_PSEME STANDARD; PRT; 83 AA.
 AC Q00457;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Toluene-4-monooxygenase system protein B (EC 1.14.13.-).
 GN TMOB.
 OS Pseudomonas mendocina.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=300;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.
 RC STRAIN=KRI;
 RX MEDLINE=91358306; PubMed=1885512;
 RA Yen K.-M., Karl M.R., Blatt L.M., Simon M.J., Winter R.B.,
 RA Fausset P.R., Lu H.S., Harcourt A.A., Chen K.K.;
 RT "Cloning and characterization of a Pseudomonas mendocina KRI gene
 RT cluster encoding toluene-4-monooxygenase."
 RL J. Bacteriol. 173:5315-5327(1991).
 CC -!- FUNCTION: HYDROXYLATES TOLUENE TO FORM P-CRESOL.
 CC -!- COFACTOR: FAD; requires Fe(2+) for activity.
 CC -!- PATHWAY: Toluene degradation; first step.
 CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME TOLUENE-4-MONOOXYGENASE
 CC IS FORMED BY THE TMOA, TMOB, TMOC, TMOD, TMOE AND TMOF
 CC POLYPEPTIDES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M65106; AAA26000.1; -.
 KW Aromatic hydrocarbons catabolism; Oxidoreductase; Flavoprotein;
 KW Monooxygenase; FAD; Iron.
 FT INIT MET 0 0
 SQ SEQUENCE 83 AA; 9457 MW; 4729FEF73F266F44 CRC64;

Query Match 9.9%; Score 42; DB 1; Length 83;
 Best Local Similarity 26.5%; Pred. No. 5e+02;
 Matches 13; Conservative 9; Mismatches 19; Indels 8; Gaps 2;

Qy 6 CSSQSISP-----MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEE 48
 | :: ::| :| | : :| : ||| : | ||
 Db 37 CVNRRVAPREGVMVRVKHRSTELFPRDMTIAESGL--NPTEVIDVVFE 83

RESULT 15

HFOB_METFO

ID HFOB_METFO STANDARD; PRT; 67 AA.
AC P48784;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Archaeal histone B.
GN HFOB.
OS Methanobacterium formicicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanobacterium.
OX NCBI_TaxID=2162;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JF-1;
RX MEDLINE=95138058; PubMed=7836329;
RA Darcy T.J., Sandman K.M., Reeve J.N.;
RT "Methanobacterium formicicum, a mesophilic methanogen, contains three
RT Hfo histones.";
RL J. Bacteriol. 177:858-860(1995).
CC -!- FUNCTION: Binds and compact DNA (95 to 150 base pairs) to form
CC nucleosome-like structures that contain positive DNA supercoils.
CC -!- SUBUNIT: Homodimer or heterodimer (Potential).
CC -!- SIMILARITY: Belongs to the archaeal histone HMF family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12929; AAA67720.1; -.
DR HSSP; P48781; 1B67.
DR InterPro; IPR003958; CBFA_NFYB_domain.
DR InterPro; IPR007124; Hist_TAF.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
KW DNA-binding; Multigene family.
SQ SEQUENCE 67 AA; 7149 MW; 1132F83ACAD88445 CRC64;

Query Match 9.8%; Score 41.5; DB 1; Length 67;
Best Local Similarity 26.9%; Pred. No. 4.4e+02;
Matches 21; Conservative 14; Mismatches 20; Indels 23; Gaps 5;

Qy 11 ISPMRSISENSILVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKKG-----CLRLGTH- 63
|:|: | :|: || :: |||: |:| | | ::| |
Db 5 IAPIGRIIKNA-----GAERVSDDAREALAKALEE-----KGETIATEAVKLAKHA 50
Qy 64 GSPTASSQSSATNMAIHR 81
| | :| :|: |
Db 51 GRKTV--KASDVELAVKR 66

RESULT 16

RPON_SULSO

ID RPON_SULSO STANDARD; PRT; 66 AA.
AC Q980Z8;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).
 GN RPON OR SSO5140.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
 CC of DNA into RNA using the four ribonucleoside triphosphates as
 CC substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE S.ACIDOCALDIARUS RNAP IS COMPOSED OF 13 SUBUNITS.
 CC -!- SIMILARITY: Belongs to the archaeal rpoN / eukaryotic RPB10 RNA
 CC polymerase subunit family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE006647; AAK40429.1; -.
 DR PIR; F90146; F90146.
 DR HAMAP; MF_00250; -, 1.
 DR InterPro; IPR000268; RNA_pol_N.
 DR Pfam; PF01194; RNA_pol_N; 1.
 DR ProDom; PD006539; RNA_pol_N; 1.
 DR PROSITE; PS01112; RNA_POL_N_8KD; 1.
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Metal-binding; Complete proteome.
 FT METAL 7 7 ZINC (BY SIMILARITY).
 FT METAL 10 10 ZINC (BY SIMILARITY).
 FT METAL 44 44 ZINC (BY SIMILARITY).
 FT METAL 45 45 ZINC (BY SIMILARITY).
 SQ SEQUENCE 66 AA; 7591 MW; C6774B541A1CFA13 CRC64;

Query Match 9.7%; Score 41; DB 1; Length 66;
 Best Local Similarity 25.7%; Pred. No. 5e+02;
 Matches 19; Conservative 12; Mismatches 29; Indels 14; Gaps 3;

```

Qy      13 PMRSISENSLVAMDFSGQKSRVI--ENPTEALSVAVEEGLAWRKKGCLRLGTHGSPTASS 70
      || : ||| : : || ||| : | : | :: | | : |
Db      4 PIRCFTCGSLIADKWQSFITRVNAGENPGKVL-----DDLGVKRYCCRRM-----LLS 51

Qy      71 QSSATNMAIHRSQP 84
      | || ::|
Db      52 HVDIINEVIHYTRP 65

```

RESULT 17

DC13 HUMAN

ID DC13 HUMAN STANDARD; PRT; 79 AA.

AC 09NR^P2;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE UPF0287 protein DC13.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

```
RC TISSUE=Dendritic cell;
```

RA Gu Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.;

RT "Novel genes expressed in human dendritic cells.";

RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schee

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Sm

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

BT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

CC -!- SIMILARITY: Belongs to the UPF0287 family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF201935; AAF86871.1; -.
DR EMBL; BC032631; AAH32631.1; -.
SQ SEQUENCE 79 AA; 9460 MW; 783381BD6DAFB7AA CRC64;

Query Match 9.7%; Score 41; DB 1; Length 79;
Best Local Similarity 40.0%; Pred. No. 6.2e+02;
Matches 10; Conservative 5; Mismatches 6; Indels 4; Gaps 1;

Qy 31 KSRVIENPTEALSVAVEEGLAWRKK 55
|: :|||:: | |:| |||
Db 49 KNEYVENRTKSR----EHGIAMRKK 69

RESULT 18

CC3_CARCN

ID CC3_CARCN STANDARD; PRT; 43 AA.
AC P32956;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cysteine proteinase III (EC 3.4.22.-) (CC-III) (Fragment).
OS Carica candamarcensis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=29731;
RN [1]
RP SEQUENCE.
RC TISSUE=Latex;
RX MEDLINE=94030669; PubMed=8216902;
RA Walreavens V., Jaziri M., van Beeumen J., Schnek A.G.,
RA Kleinschmidt T., Looze Y.;
RT "Isolation and preliminary characterization of the cysteine-
RT proteinases from the latex of Carica candamarcensis Hook.";
RL Biol. Chem. Hoppe-Seyler 374:501-506(1993).
CC -!- PTM: Glycosylated.
CC -!- SIMILARITY: Belongs to peptidase family C1.
DR HSSP; P14080; 1YAL.
DR MEROPS; C01.020; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
KW Hydrolase; Thiol protease; Glycoprotein.
FT ACT_SITE 25 25 BY SIMILARITY.
FT NON_TER 43 43
SQ SEQUENCE 43 AA; 4636 MW; F4C5D2881886E291 CRC64;

Query Match 9.6%; Score 40.5; DB 1; Length 43;
Best Local Similarity 32.5%; Pred. No. 3.3e+02;

Matches 13; Conservative 5; Mismatches 15; Indels 7; Gaps 2;

```
Qy      48 EGLAWRKKGCL----RLGTHGSPTASSQSSAT---NMAIH 80
      | : |||| :      | : || | | :      | : |
Db      3 ESIDWRKKGAVTPVKNQGSCGSCWAFSTIATVEGINKIVH 42
```

RESULT 19

GBG5_HUMAN

ID GBG5_HUMAN STANDARD; PRT; 68 AA.
AC P30670; Q61015;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-5 subunit.
GN GNG5 OR GNGT5.
OS Homo sapiens (Human),
OS Mus musculus (Mouse),
OS Rattus norvegicus (Rat), and
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606, 10090, 10116, 9913;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RX MEDLINE=99009227; PubMed=9790912;
RA Liu B., Aronson N.N. Jr.;
RT "Structure of human G protein Ggamma5 gene GNG5.";
RL Biochem. Biophys. Res. Commun. 251:88-94(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Human; TISSUE=Blood;
RX MEDLINE=98318631; PubMed=9653160;
RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
RA Wang Y.-X., Chen S.-J., Chen Z.;
RT "Identification of genes expressed in human CD34(+) hematopoietic
RT stem/progenitor cells by expressed sequence tags and efficient full-
RT length cDNA cloning.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RA Puhl H.L. III, Ikeda S.R., Aronstam R.S.;
RT "cDNA clones of human proteins involved in signal transduction
RT sequenced by the Guthrie cDNA resource center (www.cdna.org).";
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Human; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]

RP SEQUENCE FROM N.A.

RC SPECIES=Bovine, and Rat; TISSUE=Liver;
 RX MEDLINE=92195304; PubMed=1549114;
 RA Fisher K.J., Aronson N.N. Jr.;
 RT "Characterization of the cDNA and genomic sequence of a G protein
 RT gamma subunit (gamma 5).";
 RL Mol. Cell. Biol. 12:1585-1591(1992).
 RN [6]

RP SEQUENCE.

RC SPECIES=Bovine; TISSUE=Spleen;
 RX MEDLINE=93356792; PubMed=8352779;
 RA Morishita R., Masuda K., Niwa M., Kato K., Asano T.;
 RT "Identification of three forms of the gamma subunit of G proteins
 RT isolated from bovine spleen.";
 RL Biochem. Biophys. Res. Commun. 194:1221-1227(1993).
 RN [7]

RP SEQUENCE OF 8-53 FROM N.A.

RC SPECIES=Mouse; STRAIN=CF-1 / Harlan;
 RX MEDLINE=97011591; PubMed=8858601;
 RA Williams C.J., Schultz R.M., Kopf G.S.;
 RT "G protein gene expression during mouse oocyte growth and maturation,
 RT and preimplantation embryo development.";
 RL Mol. Reprod. Dev. 44:315-323(1996).

CC -!- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are
 CC involved as a modulator or transducer in various transmembrane
 CC signaling systems. The beta and gamma chains are required for the
 CC GTPase activity, for replacement of GDP by GTP, and for G protein-
 CC effector interaction.

CC -!- SUBUNIT: G proteins are composed of 3 units, alpha, beta and
 CC gamma.

CC -!- TISSUE SPECIFICITY: Expressed in a variety of tissues.

CC -!- SIMILARITY: Belongs to the G protein gamma family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC

DR EMBL; AF085709; AAC72203.1; -.
 DR EMBL; AF085708; AAC72203.1; JOINED.
 DR EMBL; AF038955; AAC39869.1; -.
 DR EMBL; AF493873; AAM12587.1; -.
 DR EMBL; BC003563; AAH03563.1; -.
 DR EMBL; M95779; AAA30535.1; -.
 DR EMBL; M95780; AAA41188.1; -.
 DR EMBL; U38498; AAB01729.1; -.
 DR PIR; B42243; B42243.
 DR Genew; HGNC:4408; GNG5.
 DR MIM; 600874; -.
 DR MGD; MGI:109164; Gng5.
 DR InterPro; IPR001770; G-gamma.
 DR Pfam; PF00631; G-gamma; 1.
 DR PRINTS; PR00321; GPROTEING.
 DR ProDom; PD003783; G-gamma; 1.
 DR SMART; SM00224; GGL; 1.
 DR PROSITE; PS50058; G_PROTEIN_GAMMA; 1.
 KW Transducer; Prenylation; Lipoprotein; Multigene family.
 FT LIPID 65 65 S-geranylgeranyl cysteine
 FT (By similarity).
 FT PROPEP 66 68 REMOVED IN MATURE FORM (BY SIMILARITY).
 SQ SEQUENCE 68 AA; 7318 MW; 9AF7A16558863602 CRC64;

Query Match 9.6%; Score 40.5; DB 1; Length 68;
 Best Local Similarity 36.4%; Pred. No. 5.8e+02;
 Matches 12; Conservative 5; Mismatches 15; Indels 1; Gaps 1;

QY 44 VAVEEGLAWRKKGCLRLGTHGSPTASSQSSATN 76
 | | : | | : | | : | | :
 Db 26 VKVSQAADLKQFCLQNAQH-DPLLTGVSSSTN 57

RESULT 20

BAXC_HUMAN

ID BAXC_HUMAN STANDARD; PRT; 41 AA.
 AC Q07815;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE BAX protein, cytoplasmic isoform gamma.
 GN BAX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=B-cell;
 RX MEDLINE=93364978; PubMed=8358790;
 RA Oltvai Z.N., Milliman C.L., Korsmeyer S.J.;
 RT "Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that
 RT accelerates programmed cell death."
 RL Cell 74:609-619(1993).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;

```

CC      Name=Alpha;
CC      IsoId=Q07815-1; Sequence=Displayed;
CC      Name=Beta;
CC      IsoId=Q07815-2; Sequence=Not described;
CC      Name=Gamma;
CC      IsoId=Q07815-4; Sequence=Not described;
CC      Name=Delta;
CC      IsoId=Q07815-3; Sequence=Not described;
CC      -!- SIMILARITY: Belongs to the Bcl-2 family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L22475; AAA03621.1; -.
DR      PIR; C47538; C47538.
DR      Genew; HGNC:959; BAX.
DR      MIM; 600040; -.
DR      GO; GO:0008637; P:apoptotic mitochondrial changes; TAS.
DR      GO; GO:0007281; P:germ-cell development; TAS.
DR      GO; GO:0008624; P:induction of apoptosis by extracellular sig. . .; TAS.
DR      GO; GO:0008634; P:negative regulation of survival gene products; TAS.
KW      Apoptosis; Alternative splicing.
SQ      SEQUENCE 41 AA; 4678 MW; D94639AABB927859 CRC64;

```

```

Query Match          9.5%; Score 40; DB 1; Length 41;
Best Local Similarity 31.0%; Pred. No. 3.6e+02;
Matches 13; Conservative 5; Mismatches 12; Indels 12; Gaps 2;

```

```

Qy      25 MDFSGQKSRVIENPTEALSVAVEEG-----LAWRKKGCLRL 60
      || ||      | |      :|      :|:|      :|      |||:
Db      1 MDGSG-----EQPRGGVSSRIEQGEWGGRHPSWPWTRCLRM 36

```

RESULT 21

HS2M_LYCES

```

ID      HS2M_LYCES      STANDARD;      PRT;      56 AA.
AC      P81161;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Heat shock 22 kDa protein, mitochondrial (Fragments).
OS      Lycopersicon esculentum (Tomato).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC      lamiids; Solanales; Solanaceae; Solanum.
OX      NCBI_TaxID=4081;
RN      [1]
RP      SEQUENCE.
RC      STRAIN=cv. Sweet;
RX      MEDLINE=98345975; PubMed=9680997;
RA      Banzet N., Richaud C., Deveaux Y., Kazmaier M., Gagnon J.,
RA      Triantaphylides C.;

```

RT "Accumulation of small heat shock proteins, including mitochondrial
RT HSP22, induced by oxidative stress and adaptive response in tomato
RT cells.";
RL Plant J. 13:519-527(1998).
CC -!- FUNCTION: May play a protective role against oxidative stress.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- INDUCTION: By heat shock, and under other conditions of stress,
CC such as increased salt concentration and starvation.
CC -!- SIMILARITY: Belongs to the small heat shock protein (HSP20)
CC family.
DR InterPro; IPR002068; Hsp20.
DR PROSITE; PS01031; HSP20; PARTIAL.
KW Heat shock; Mitochondrion.
FT NON_CONS 14 15
FT UNSURE 15 15
FT NON_CONS 35 36
FT UNSURE 36 36
FT NON_TER 56 56
SQ SEQUENCE 56 AA; 6446 MW; 2AB9F927C7720076 CRC64;

Query Match 9.5%; Score 40; DB 1; Length 56;
Best Local Similarity 39.1%; Pred. No. 5.3e+02;
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 38 PTEALSVAVEEGLAWRKKGCLRL 60
| | : ||:| | | | :
Db 21 PVENVRVALEENTLIMKNGVLKV 43

RESULT 22

RPON_THEAC

ID RPON_THEAC STANDARD; PRT; 72 AA.
AC Q9HL09;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).
GN RPON OR TA0431.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SIMILARITY: Belongs to the archaeal rpoN / eukaryotic RPB10 RNA

CC polymerase subunit family.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL445064; CAC11573.1; -.
DR HSSP; O26147; 1EF4.
DR HAMAP; MF_00250; -; 1.
DR InterPro; IPR000268; RNA_pol_N.
DR Pfam; PF01194; RNA_pol_N; 1.
DR ProDom; PD006539; RNA_pol_N; 1.
DR PROSITE; PS01112; RNA_POL_N_8KD; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KW Metal-binding; Complete proteome.
FT METAL 7 7 ZINC (BY SIMILARITY).
FT METAL 10 10 ZINC (BY SIMILARITY).
FT METAL 53 53 ZINC (BY SIMILARITY).
FT METAL 54 54 ZINC (BY SIMILARITY).
SQ SEQUENCE 72 AA; 8368 MW; 792AEDA20E5447E2 CRC64;

Query Match 9.5%; Score 40; DB 1; Length 72;
Best Local Similarity 31.6%; Pred. No. 7.1e+02;
Matches 12; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy 11 ISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEE 48
| |:| | ::| |: || | | : ||
Db 2 IIPVRCFSCGRVIASDYGRYIKRVNEIKAEGRDPSPEE 39

RESULT 23

RPON_THEVO

ID RPON_THEVO STANDARD; PRT; 72 AA.
AC Q979K0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).
GN RPON OR TV1161 OR TVG1188103.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).

CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
 CC of DNA into RNA using the four ribonucleoside triphosphates as
 CC substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- SIMILARITY: Belongs to the archaeal rpoN / eukaryotic RPB10 RNA
 CC polymerase subunit family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP000995; BAB60303.1; -.
 DR HAMAP; MF_00250; -; 1.
 DR InterPro; IPR000268; RNA_pol_N.
 DR Pfam; PF01194; RNA_pol_N; 1.
 DR ProDom; PD006539; RNA_pol_N; 1.
 DR PROSITE; PS01112; RNA_POL_N_8KD; 1.
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Metal-binding; Complete proteome.
 FT METAL 7 7 ZINC (BY SIMILARITY).
 FT METAL 10 10 ZINC (BY SIMILARITY).
 FT METAL 53 53 ZINC (BY SIMILARITY).
 FT METAL 54 54 ZINC (BY SIMILARITY).
 SQ SEQUENCE 72 AA; 8483 MW; 06AEC0AA7AC75CA6 CRC64;

Query Match 9.5%; Score 40; DB 1; Length 72;
 Best Local Similarity 28.9%; Pred. No. 7.1e+02;
 Matches 11; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

Qy 11 ISPMRSISENSILVAMDFSGQKSRVIENPTEALSVAVEE 48
 | | : | | : | : | : |
 Db 2 IIPVRCFSCGRVIASDYGRYLRRINEIRSEGREPTAEE 39

RESULT 24

PBP_HYACE

ID PBP_HYACE STANDARD; PRT; 35 AA.
 AC P34175;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pheromone-binding protein (PBP) (Fragment).
 OS Hyalophora cecropia (Cecropia moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
 OC Saturniidae; Saturniinae; Attacini; Hyalophora.
 OX NCBI_TaxID=7123;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=91186129; PubMed=2010751;
 RA Vogt R.G., Prestwich G.D., Lerner M.R.;
 RT "Odorant-binding-protein subfamilies associate with distinct classes

RT of olfactory receptor neurons in insects.";
 RL J. Neurobiol. 22:74-84(1991).
 CC -!- FUNCTION: THIS MAJOR SOLUBLE PROTEIN IN OLFACTORY SENSILLA OF MALE
 CC MOTHS MIGHT SERVE TO SOLUBILIZE THE EXTREMELY HYDROPHOBIC
 CC PHEROMONE MOLECULES AND TO TRANSPORT PHEROMONE THROUGH THE AQUEOUS
 CC LYMPH TO RECEPTORS LOCATED ON OLFACTORY CILIA.
 CC -!- TISSUE SPECIFICITY: Antenna.
 CC -!- SIMILARITY: Belongs to the PBP/GOBP family.
 DR HSSP; P34174; 1DQE.
 DR InterPro; IPR006170; PBP_GOBP.
 DR Pfam; PF01395; PBP_GOBP; 1.
 KW Pheromone-binding; Pheromone response; Transport.
 FT NON_TER 35 35
 SQ SEQUENCE 35 AA; 4061 MW; 9B1B9D20D472E769 CRC64;

Query Match 9.3%; Score 39.5; DB 1; Length 35;
 Best Local Similarity 37.9%; Pred. No. 3.4e+02;
 Matches 11; Conservative 5; Mismatches 10; Indels 3; Gaps 1;

QY 14 MRSISENSLVAMDFSGQKSRVIENPTEAL 42
 |:|:| | | | : : | :
 Db 5 MKSLSENFCKAMD---QCKQELNLPDEVI 30

RESULT 25

Y574_LACLA

ID Y574_LACLA STANDARD; PRT; 60 AA.
 AC Q9CHZ4;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable tautomerase LL0574 (EC 5.3.2.-).
 GN LL0574.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis ssp. lactis IL1403."
 RL Genome Res. 11:731-753(2001).
 CC -!- SIMILARITY: Belongs to the tautomerase family.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AE006291; AAK04672.1; -.
 DR PIR; F86696; F86696.

DR HAMAP; MF_00718; -: 1.
 DR InterPro; IPR004370; Taut.
 DR Pfam; PF01361; Tautomerase; 1.
 DR ProDom; PD404143; Taut; 1.
 KW Isomerase; Complete proteome.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT ACT_SITE 1 1 CATALYTIC BASE (BY SIMILARITY).
 SQ SEQUENCE 60 AA; 6667 MW; 19E80C7BA3EAFFFF CRC64;

Query Match 9.3%; Score 39.5; DB 1; Length 60;
 Best Local Similarity 21.4%; Pred. No. 6.5e+02;
 Matches 9; Conservative 14; Mismatches 16; Indels 3; Gaps 1;

Qy 15 RSISENSLVAMDFSGQKSRVIENPTEALSVA---VEEGLAWR 53
 |:: : :::| : : |: || |: | : ||: ::
 Db 11 RTVEQKAIIAKEITESISKHAGAPTSIAHVIFNDLPEGMLYQ 52

Search completed: July 8, 2004, 08:20:07
 Job time : 28.4567 secs